

Industrial 'R' up in a downturn

In markets that have never been tougher, IT and telecoms companies are well structured for essential innovation. Investors are recognizing this, but some governments risk losing long-term opportunities by failing to invest in related research.

Even before last month's terrorist attacks in the United States pitched the world's economies downward, the computing and telecoms industries had been badly hit by stagnant markets and colossal debts. More than ever, now, these companies have to be at the forefront of technological innovation if they are to remain in the game.

Since the previous recessions, the rate of introduction of new products has increased vastly. What was previously considered 'long-term' research and development on a ten-year horizon now often needs to be pushed through in two years. This is markedly changing the way companies view their R&D investments. A defensive strategy of cutting R&D and seeking to prop up one's current products is now widely acknowledged as a recipe for corporate suicide. For larger companies that are not already peering over the edge of the abyss, the trend is to protect and indeed increase R&D budgets despite the downturn (see News Feature, page 448).

Industrial 'R' — basic research, defined as research unconnected with current products — typically represents just 10% of a company's total R&D budget. This relatively small investment, whether intra- or extramural, is the lifeblood of new generations of products. Where more shake-ups will occur is in the 'D'. In a downturn, managers will be looking harder at which projects to reinforce or cut. Investors watch carefully for companies that seem to be intelligently directing R&D funding towards higher growth areas.

Keys to survival

Whether companies can maintain the scale of their effort if recession bites is the big question. In September 2000, the stocks of Sun Microsystems, a pillar of the high-tech industry, closed at US\$61.81 a share, and it reported quarter profits of \$510 million. This week its stock stood at \$8.27 a share, and it is predicted to post quarter losses of more than \$210 million. Sun is making swingeing cuts across the board — except in R&D. Encouragingly, industry analysts are backing the company, advising that it has emerged reinforced from previous recessions and is a sound long-term bet. Some predict a return to profit by early 2002. Educating investors in the long-term value of stocks of research-based companies, and openness by companies on the how and why of the hard choices being made, will be key to maintaining investor confidence and preventing a flight from the sector.

Small companies with unproven products in unproven markets are particularly vulnerable to becoming victims of this shake-out. But the venture-capital industry is much larger and has greater expertise than a decade ago, and although funds are more difficult to obtain, it will continue to support the most promising companies. Other companies may survive by being bankrolled by larger companies on the lookout for promising technologies.

Today's high-tech giants also know that in this phase of the cycle, they risk being left stranded by companies that innovate and adapt faster. Intel, for example, has increased R&D this year from \$3.9 billion to \$4.2 billion. Its multipurpose Pentium processors risk becoming commodities, and it is turning to research on chips that will create a world where hundreds of computers sense their environment

and interact with one another and their users. The availability of such chips could fuel other companies to kick-start a new economy based on pervasive computing devices — with, Intel hopes, 'Intel Inside'.

IBM likewise sees opportunity amid the gloom. Businesses worldwide are seeking to cut costs because of the downturn, but the major costs of IT are not the hardware but maintenance and administration. IBM is therefore seeking to create self-healing and self-managing computers. Although this will involve much basic research, the 'D' will also be essential. The computer industry knew for decades that copper interconnects would make for better chips, but it took IBM years of 'D' to manufacture the first copper chips in 1998.

Changing approach

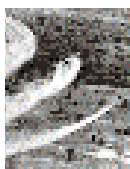
For companies in this sector that survive the potentially devastating economic climate, research is likely to prove resilient, thanks, above all, to the shake-up of the management and prioritizing of R&D over the past decade or so. That sturdiness is highlighted in a study of six large companies due to be published this week (<http://les1.man.ac.uk/PREST/TSF>). It testifies to a series of changes that add up to a well-tuned corporate R&D culture. Multidisciplinary and multinational approaches to strategic technologies are commonplace. Despite rumours to the contrary, there is still a recognition of the value of the culture of basic science.

On the other hand, business units and central R&D have become more integrated in project development, reflecting a recognition that technology goals can stimulate fundamental research and vice versa. Interactions with the public science base have also evolved considerably over the past decade, and companies are now more adept at prospecting the best basic research laboratories worldwide to identify and develop winners.

This throws the spotlight onto the strength of public funding of basic research in related disciplines. A series of 'innovation scoreboards' published over the past four weeks alone, by the Organisation for Economic Co-operation and Development, the European Commission and the UK Department of Trade and Industry, provide comprehensive data on a wide range of factors influencing innovation for countries and firms, from investment in R&D to the supply of science and engineering graduates, and analysis of industrial/academic structures. The data should be interpreted cautiously, but some of the composite snapshots should set alarm bells ringing. In Europe, for example, Ireland, Denmark, Finland and Sweden are 'moving ahead', whereas the United Kingdom and Germany are 'losing momentum' and France and Italy are 'falling further behind'.

Innovation in the information industries has come a long way in the past ten years. It is now a complex partnership between large companies, small companies, a solid venture-capital market, and public research. Investors appreciate the progress made. But countries that see IT and telecoms as important sectors for their economies should beware of failing to attract and stimulate the industries that have worked hard to adapt and are better able to move their operations to wherever well-trained talent and a solid science base are to be found. ■





Dioxin damage
US and Vietnam join
forces to assess effects
of Agent Orange
p442



Test case
Europe steps up
opposition to
cancer-gene patent
p443



Full mouse
Physical map of
mouse genome
unveiled
p444



Black hole
Partners nervous
as US ponders
space-station cuts
p445

Large Hadron Collider in crisis as magnet costs spiral upwards

David Adam, London

Europe's attempt to build the world's most powerful particle accelerator was plunged into crisis this week as project managers admitted that it faces cost overruns of several hundred million dollars.

CERN, the European particle physics laboratory, where the Large Hadron Collider (LHC) is being built, will face years of budget cuts and belt-tightening as a result of the overruns, its management told staff. But this will cover only a fraction of the extra costs, and CERN is preparing to take the awkward step of asking its 20 member states to cover the rest.

Scheduled to open in 2006, the LHC is intended to be the world's next major high-energy physics facility. But the technical challenge of designing and building the superconducting magnets needed to steer protons and ions through the LHC's accelerators has proven more difficult — and expensive — than CERN expected. Problems with such magnets contributed to the 1993 demise of the US Superconducting Supercollider.

CERN says that the higher costs of the magnets, together with greater installation and civil engineering costs, have put the core project some SFr480 million (US\$300 million) over its original SFr2.6 billion budget. But scientists at CERN who have been briefed on the situation say it could cost at least SFr3.4 billion — SFr800 million more than initially planned — to get the LHC up and running. Prototype magnets already built have cost the laboratory about SFr150 million more than it expected, and some SFr220 million extra must also be found to install detectors and to pay for computers.

CERN now has until 7 November to come up with a rescue plan to present to its finance committee, made up of representatives from the European countries funding the facility. "The organization is reviewing various possible options including cuts in the scientific programme, reductions in spending within the organization, bank loans and extra contributions from member states," says Neil Calder, a spokesman for CERN.

At a crisis meeting of staff and users on 1 October, CERN director-general Luciano



Running short: CERN may turn to its member states for the cash to finish the Large Hadron Collider.

Maiani said that the organization was aiming for 10% cost reductions across all divisions. "Basically we're looking at austerity measures over the next few years to pick up the slack," says one CERN researcher.

Some researchers expressed unhappiness that they first learned of the problems from an article in a local newspaper. Their irritation is shared by the CERN member govern-

ments who are paying for the project.

A spokesman for the Particle Physics and Astronomy Research Council, which pays Britain's annual £65-million (US\$96-million) subscription to CERN, says that the situation is "potentially very serious". But he says he will wait for the November meeting before commenting on the possible implications for British physics and astronomy. ■

Senators call for biodefence boost

Jonathan Knight

A massive increase in public-health spending is needed to prepare the United States for a possible terrorist attack with biological weapons, two prominent US senators say.

In response to the attacks on the World Trade Center and the Pentagon on 11 September, Senators Edward Kennedy (Democrat, Massachusetts) and Bill Frist (Republican, Tennessee) say that the US government should spend \$1.6 billion next year to bolster public health and biodefence.

According to a report from the General Accounting Office, federal spending on bioterrorism readiness, including research, amounted to about \$350 million in the 2001 fiscal year, which ended on 30 September.

Although this is a significant increase from previous years, biodefence efforts are

poorly coordinated, the report says. For example, several agencies disagree over which pathogens should count as potential threats. In addition, most hospitals are poorly prepared to detect or handle a massive outbreak of disease.

To remedy this, the Kennedy–Frist proposal includes \$625 million for state and local health agencies — a sixfold increase.

"There does seem to finally be an acknowledgement that the public-health infrastructure is in deep disarray," says Alan Zelicoff, a biodefence expert at Sandia National Laboratory in New Mexico.

Mark Wheelis, a microbiology professor at the University of California, Davis, says that the large number of ways an attack could be perpetrated will make it hard to decide how best to spend the money. ■

Bilateral Vietnam study plans to assess war fallout of dioxin

Rex Dalton, San Diego

Researchers from the United States and Vietnam met this week in Hawaii to finalize field-testing methods for the nations' first bilateral research project on dioxin pollution from the defoliant Agent Orange.

The scientists plan to begin sampling soil throughout Vietnam this winter to create

a database for research on the health and environmental impacts of dioxins spread by the US military during the Vietnam War.

"This research programme can be a benefit to the whole world," says Christopher Portier, director of environmental toxicology at the US National Institute of Environmental Health Sciences (NIEHS), which is planning the project. "But it first must be beneficial to the Vietnamese."

An international conference on dioxin and other herbicides is set to take place in Hanoi in March. Preliminary results from initial testing may be available for that meeting, officials say.

More than 30 years ago, the US military used Agent Orange as a defoliant on vast stretches of Vietnamese jungle. Aeroplane spraying, along with dioxin dumping and storage facilities, left many 'hot spots' where chemical pollution is thought to be extensive.

Individual researchers from the United States, Canada and other countries have found that dioxin residue has polluted food supplies, including fish. Subsequent studies have shown that those eating the fish have high levels of dioxins, which can cause cancer, immune disorders and birth defects.

During the past two years, the NIEHS and the US Environmental Protection Agency have worked to develop a \$400,000 research project designed to provide a comprehensive overview of dioxin in Vietnam (see *Nature* 406, 818; 2000).

Eviction threat to Belgium's science academies

Quirin Schiermeier, Munich

Belgian scientists are on a collision course with their government over plans to evict the country's science academies from their home in central Brussels.

The Palace of the Academies, built in the 1820s and considered an outstanding example of the architecture of the time, houses the Royal Flemish Academy of Belgium for Science and the Arts (KVAB) and its French-speaking equivalent, the Royal Academy of Science, Literature and Arts of Belgium.

Members of the academies first discovered the government's plan to take back their building in July. "We got the news from a television report," says Niceas Schamp, an organic chemist at Ghent University and permanent secretary of the KVAB. "There was no warning, nobody was informed and no consultations were held with the management of the academies."

The KVAB is currently attempting to convince the government to abandon its plan. A statement drafted last month, calling on Patrick Dewael, prime minister of the regional government of Flanders, to oppose the move, has received 900 signatures, mainly from academy members.

Schamp says the threatened move would be a "humiliating loss of status" to his organization. The palace currently serves as a meeting and working place for eminent scientists and artists.



Deadly action: spraying with Agent Orange some 30 years ago left many 'hot spots' of pollution.

Doctors propose panel on research misconduct

Erica Klarreich, London

British physicians are proposing the formation of a national panel to handle investigations of misconduct in biomedical research.

The panel — suggested by doctors' organizations in the absence of any government plan for such a body — would also provide research institutions with guidelines for good scientific practice and advice for dealing with misconduct allegations.

The plan would tackle what its proposers see as a serious anomaly in current investigations of biomedical-misconduct investigations. Currently, physicians engaged in research are subject to investigation by the General Medical Council, whereas scientists conducting similar work are not monitored by any national body.

The proposal stems from a conference in

November 1999 involving the Royal physicians' colleges of Edinburgh, Glasgow and London, and their joint Faculty of Pharmaceutical Medicine. The recommendations will be discussed at a London meeting of COPE, an organization of biomedical-journal editors, on 15 October.

"We felt that there were sufficient instances of institutions having problems handling this locally that they need a larger body to turn to for advice and perhaps external inspections," says Gordon Lowe of the Royal College of Physicians of Edinburgh, who helped to draw up the plans.

The physicians hope that the panel will include members from academic, medical and industrial backgrounds. Its main focus would be to help institutions to set their own guidelines and conduct investigations, Lowe says, but it might also investigate certain

cases itself. The proposal's backers want government funding to support the panel, and plan to hold talks with government officials in Edinburgh and London to seek such support.

Not everyone is happy with the idea of a UK national body to oversee only biomedical research, however. "We need a central government body that integrates all the sciences and engineering, instead of having bits and pieces done in an ad hoc manner by various bodies," says Raymond Spier, an ethicist at the University of Surrey.

A recent Wellcome Trust plan for handling scientific misconduct was hailed by some observers as a unifying force for misconduct investigations in the United Kingdom, but it will apply only to research funded by the trust (see *Nature* 412, 667; 2001).

Testing time for gene patent as Europe rebels

Meredith Wadman, Washington

A European rebellion against the patent on a gene for breast cancer held by US company Myriad Genetics is gathering pace.

Paris's Curie Institute and the nearby Gustave Roussy Institute will next week file a formal objection to the patent, which covers the diagnostic test for the *BRCA1* gene (see *Nature* **413**, 95–96; 2001). But this move is just one aspect of the mounting opposition the patent faces across Europe.

Despite Myriad's request that all test samples be sent to its labs in Salt Lake City, scores of geneticists and dozens of diagnostic labs throughout Europe are still doing their own *BRCA1* testing. They say they are reluctant to concede their independence to a firm whose costs substantially exceed their own.

Bernhard Weber, head of breast cancer genetic testing at the Institute of Human Genetics in Würzburg — one of 12 non-profit labs that between them carry out the bulk of genetic testing in Germany — says that, to his knowledge, not one of the labs has begun sending samples to Myriad. "There was a resentment in the group, in all 12 centres, saying 'We do not want to support this patent,'" says Weber. "We need to get our own results based on our own technology."

In the Netherlands, the network of eight labs that conducts all of the country's genetic testing is actively seeking legal avenues to block the Myriad patent. "We have developed a test in our own labs. It works. It is not more expensive than Myriad's. And now we should start paying royalties to Myriad?" says Bert Bakker, a molecular geneticist who heads the diagnostic lab in Leiden.

In Britain, National Health Service laboratories have continued their own *BRCA1* testing, while the health department negotiates with Myriad over the extent to which they will have to submit samples. Although British geneticists say they will abide by any negotiated agreement, their anger at Myriad runs deep. "Myriad wants to enforce a monopoly on the provision of a service. That is an unwarranted and novel restriction on medical practice," says Rob Elles, secretary of the British Society for Human Genetics, who is in charge of molecular genetic testing in the regional lab at St Mary's Hospital in Manchester.

This resistance reflects broader sentiment against the patenting of human genes in Europe, researchers say. It may be a harbinger of tough times for companies who seek to exploit gene-based discoveries there.

"The substance patents now being given to the human genome are inappropriate and endanger research and medicine. Information about the human genome can't be invented. It is the common heritage of all humans," says Otmar Kloiber, senior executive of the German Medical Association.



Hands off: Curie researchers are ignoring Myriad's patent by continuing to do their own *BRCA1* tests.

Mike Stratton, head of the cancer genome project at the Sanger Centre near Cambridge, says: "The big question is whether naturally occurring DNA sequences are really up for grabs in the way that Myriad has indicated, and whether it is possible really to implement the patents that they get."

Myriad, which discovered *BRCA1* in 1994 and has since poured tens of millions of dollars into developing its \$2,680 test, has cornered the market in the United States, where it has faced little public criticism.

Myriad claims it has every right to profit similarly in Europe. Although it has so far

taken no steps to enforce its patent, it would be "very premature" to rule out such action, says Gregory Critchfield, who heads the genetic-testing division: "We have an obligation to protect our intellectual property."

Some experts say geneticists must come to terms with a new era in which diagnostic products are increasingly expensive — and proprietary. "The knee-jerk reaction of the research and medical communities is to say 'stop the patents,'" says Rebecca Eisenberg, an expert in patent law at the University of Michigan. "That's not going to happen, so I'd like to see what their backup position is." ■

Last-minute floods sink research

Alison Abbott, Munich

Floods caused by a violent storm have badly damaged two research institutes run by the CNR, Italy's national research council, just a few weeks before half of their researchers were due to move into new premises.

The International Institute of Genetics



Trashed: the soon-to-be-vacated institute buildings after mud and water poured in.

and Biophysics (IIGB) and the Institute of Protein Biochemistry and Enzymology, both based in Naples, were hit by floods on 15 September. They were preparing to relocate to the new Integrated Centre for Biological Research in the same city.

The floods ruined electron and confocal microscopes, causing an estimated 2.5 million euros (US\$2.3 million) damage. But scientists are most upset by the loss of data and of collections of knockout and transgenic organisms including the fruitfly *Drosophila* and the worm *Caenorhabditis elegans*.

Most of the IIGB's knockout and transgenic mice were killed — only those in cages on high racks survived as water and mud poured into the animal house to a height of two metres. "Fortunately the flood hit at 5 a.m. when no one was about, and there were no casualties," says IIGB director John Guardiola.

IIGB scientists say that the 'temporary' premises — in use since 1962 — were highly susceptible to flooding. The CNR had been promising a relocation for 25 years, they say. Half the laboratories were due to move this month, with the rest following next year. ■

Mouse genome roars ahead with new map

Declan Butler

The public-domain effort to sequence the mouse genome is set to accelerate, following the completion of a physical map of the genome.

Sequence data flowing from the international Mouse Sequencing Consortium can now be anchored to the map, speeding up assembly of the complete genome. The project is the first to use human-genome data as a framework to build a map of another genome, and provides a proof-of-principle that will speed up completion of the genomes of other species.

The map, which offers a complete tiling path for each chromosome, is available at the Ensembl database (see below). In it, the 290,000 bacterial clones in the sequence have been assembled into just 554 contiguous blocks, or contigs, spanning 3,080 megabases — 98% of the entire mouse genome. This relatively small number of contigs will make it easy to assemble the full genome, researchers say.

The low number of contigs and the speed with which the map has been completed can be attributed to the usefulness of the human genome sequence as a guide, say members of the multicentre team that created the map. Mouse genes align well with human ones, and long blocks of the sequence are common to both genomes. The team used information from these common blocks to help them place and orientate the mouse-clone contigs; this enabled them to reduce the number of contigs from more than 7,000.

In turn, the mouse map is also helping to

plug remaining gaps in the human genome sequence. By aligning the mouse map next to the human sequence, mouse contigs have already been identified as spanning gaps in the human genome.

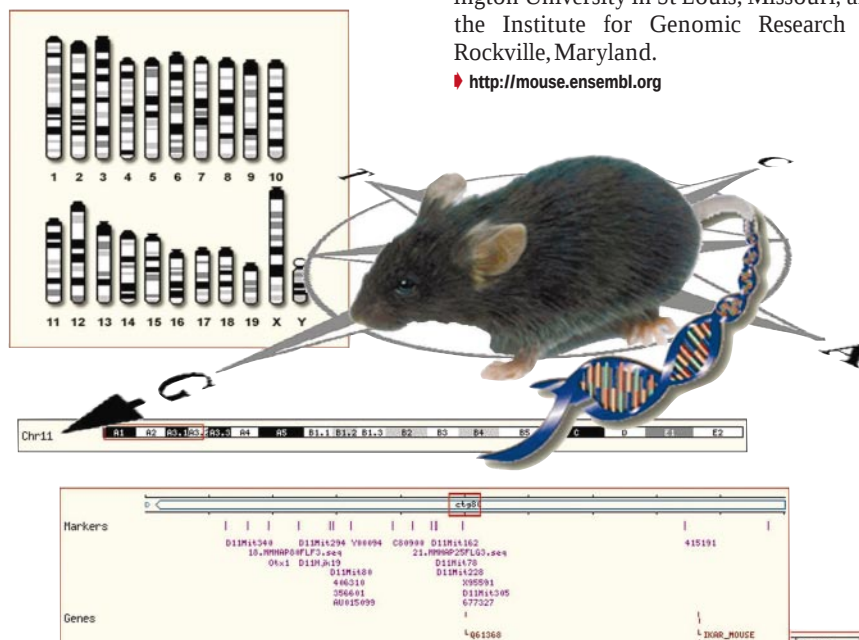
The availability of the mouse map means that assembly of the entire mouse genome can proceed apace. The Mouse Sequencing Consortium is generating sequence by applying the high-throughput 'shotgun' sequencing method pioneered by Celera Genomics of Rockville, Maryland, using

some 6 million fragments between 500 and 700 bases long (see *Nature* **411**, 121; 2001).

"We can now nail the emerging mouse assemblies to the map," says Tim Hubbard of the Sanger Centre near Cambridge in England. Hubbard expects to have produced a rough first bash at the entire genome within the next few weeks, with continual refinement over the coming months.

The map was generated by groups at the Sanger Centre; the Genome Sequence Centre in British Columbia, Canada; Washington University in St Louis, Missouri; and the Institute for Genomic Research in Rockville, Maryland.

♦ <http://mouse.ensembl.org>



Points of reference: work on the mouse genome is helping to fill gaps in the human genome sequence.

Psychiatrist launches lawsuit over 'academic freedom'

David Spurgeon, Montreal

A psychiatrist who had a job offer withdrawn by a hospital affiliated to the University of Toronto, after making critical comments about psychotropic drugs such as



David Healy: seeking damages.

Prozac, has announced that he will sue the university and the hospital for damages of Can\$9.4 million (US\$6 million).

Lawyers acting for David Healy, of the University of Wales College of Medicine, have filed a statement of claim asking Canada's Superior Court of Justice in Toronto for formal recognition of an academic's right to speak

out without fear of losing his or her position or of being reprimanded.

The lawyers say that this lawsuit is the first to be filed in Canada based on the principle of academic freedom.

Healy claims that the university and its affiliated hospital, the Centre for Addiction and Mental Health (CAMH) contravened this principle and broke their contract with him (see *Nature* **413**, 240; 2001).

The lawsuit further alleges that Healy, director of the North Wales department of psychological medicine in Bangor, was defamed as a scientist and as a physician during attempts by the hospital and the university to justify the decision to withdraw his job offer.

In September 2000, Healy accepted the post of clinical director of the mood and anxiety disorders programme at the CAMH, with an accompanying full

professorship in the university's department of psychiatry.

According to the statement of claim, the contract was rescinded in an e-mail a week after Healy gave an invited lecture at a November colloquium at the centre.

The CAMH has issued a statement saying that it stands by its decision, which was based "solely on the needs of our patients and staff".

University officials have said that Healy would still be offered an academic appointment if he obtained an appropriate medical position in one of the university's affiliated teaching hospitals.

Healy told a press conference in Toronto on 24 September that he did not intend to profit from the lawsuit. Instead, he said, he would use the proceeds — after costs and "immediate damages" — to set up a trust fund "to promote academic freedom".

NASA budget problems alarm space-station collaborators

William Triplett, Washington

Hopes are fading fast that money can be found in Washington to complete the construction of the International Space Station to anything approaching its original specification.

As international partners look on nervously, the national security crisis following the 11 September terrorist attacks on the United States seems to have eliminated any lingering chance that extra funds will be available to cover an estimated \$5-billion shortfall in the project's NASA budget (see *Nature* **412**, 465–466; 2001).

Members of Congress, including influential figures such as Senator Bill Nelson (Democrat, Florida), a former astronaut, met with project supporters on Capitol Hill last week to scramble for ways to salvage the project.

"There may not be money available but there are solutions, and that's what [the meeting] was all about," says Marc Schlather, president of ProSpace, a citizens' group that helped to organize the meeting. However, few material solutions were evident during the sombre round-table discussion of station supporters.

"We are in a moment of crisis in so many ways," said Dana Rohrabacher (Republican, California). "We need to be creative in our approaches." Rohrabacher said that the governments of Italy and Ireland might be willing to contribute \$1.5 billion and \$200 million, respectively, to the project if companies in these countries were awarded contracts of corresponding value. He also suggested that NASA place equipment contracts with as many Russian companies as possible, because they have "a tremendous capability for producing hardware at cheaper prices than anywhere else".

A former congressional staff member on space issues urged NASA to consider using commercial launch services for resupplying the space station. At present it relies on the far more expensive space-shuttle flights.

Although no final decision has been made, NASA is considering changes in the station that would reduce its crew from six or seven to three. Researchers have said that such a change would effectively end the station's usefulness as a laboratory. In a preliminary copy of a report on the station's readiness, which will be formally published early next year, the National Academy of Sciences' Space Studies Board warns that "the future of science on the International Space Station would be severely impaired" by such a change.

European and Japanese space-agency



Crew cut: scientists are concerned by proposals to reduce NASA's space-station crew to just three.

officials, meanwhile, are putting a brave face on unfolding events, while researchers with projects planned grow increasingly restive. "We are looking at all possible solutions, but we are not altogether pessimistic," says Mark Heppener, head of space-station utilization at the European Space Agency.

But another European official, who declined to be identified, said: "The scientists probably have a feeling of betrayal. We have tried very hard to select the best proposals through a new system of international peer review, and now we feel that we are all on a waiting-list; we have no idea what could happen. We hope the United States will honour the international agreements it has entered into."

Yasushi Horikawa, space-station project manager at NASDA, the Japanese space agency, says: "The worst thing would be if researchers who have been preparing projects just have to keep waiting and waiting." A NASDA spokesman added that NASA's budget problems could delay the construction and launch of Japan's Kibo research module for the station, which is due in 2004 or 2005.

• <http://www.nap.edu>

Vatican approves use of animal transplants 'to benefit humans'

Xavier Bosch

In a surprise announcement that may spur xenotransplantation research worldwide, the Pontifical Academy for Life in the Vatican has said that it does not object to the transplantation of animal organs into humans.

Although xenotransplantation may provide cells, tissues and organs to treat a variety of serious human illnesses, several groups have called for it to be banned while ethical issues and the hazards of cross-species infection are debated. In 1999, for example, the Council of Europe called for a moratorium on clinical trials involving xenotransplantation (see *Nature* **397**, 281–282; 1999).

In a detailed report released on 26 September, the academy argues that because humans enjoy a unique and superior dignity, and God has placed non-human creatures at the service of people, the sacrifice of animals is justified as long as there will be a "relevant benefit for humans". Research into transgenic animals is also "morally acceptable", the report says, because of these benefits.

The report also argues that it is irrelevant, from a theological or moral viewpoint, whether primates or non-primates, such as pigs, are used for xenotransplantation. Despite the greater immunological barrier between pigs and humans, pigs are generally favoured over primates as a source of transplant tissue for a variety of practical, ethical and safety reasons.

In a speech at the Vatican when the report was released, Marialuisa Lavitrano, a member of the Council of Europe's xenotransplantation working group, said that society should be given accurate and balanced information on the benefits and risks of xenotransplantation.





Up in smoke: the Rothera research station.

Fire destroys British Antarctic Survey research station

London The British Antarctic Survey's plans for biology research have been thrown into doubt after a fire destroyed one of the survey's field research stations.

Staff tried to extinguish the blaze at the Bonner Laboratory, a biological facility at the Rothera Research Station on the Antarctic peninsula, using a small fire engine and a snow-blowing machine, but bad weather forced them to withdraw. Nobody was hurt or put at risk by the fire, the cause of which is unknown.

The Antarctic summer is coming, and about 100 people were due to arrive at the station next month to resume their research,

which includes an analysis of Antarctic biodiversity. Accommodation at the base is unaffected, and scientists whose work does not involve the Bonner lab still plan to travel.

Radioactive-label lab marked for closure

San Diego The National Tritium Labelling Facility in California is set to close in December, after losing its annual funding of around US\$1 million from the National Institutes of Health (NIH).

The 19-year-old facility, which is located at the Lawrence Berkeley National Laboratory at the University of California, Berkeley, provides radioactive material for tagging molecules. But NIH officials say it has management problems and a weak record of assisting researchers. According to the NIH's National Center for Research Resources, which provides the funding, not enough peer-reviewed publications were produced by NIH scientists making use of the facility to justify keeping it open.

Study rejects jabs for foot-and-mouth crisis

London Vaccination programmes would have had little effect on the scale of the British foot-and-mouth epidemic,

according to one of the most thorough epidemiological studies of the outbreak yet (M. Keeling *et al. Science*, 4 October 2001 (10.1126/science.1065973)).

A team led by Matt Keeling of the University of Cambridge considered a range of vaccination strategies in the study. They predict that the most effective of these could have cut the total number of cases by 30%, but say the logistical difficulties of vaccinating large numbers of cattle would have resulted in a smaller reduction.

Keeling's study and one by Neil Ferguson and colleagues at Imperial College, London (see page 542 of this issue) conclude that a delay in introducing the policy of culling animals on farms near infected premises was largely responsible for the scale of the epidemic.

Genome institute spawns three new research centres

Washington The US National Human Genome Research Institute is to create three new interdisciplinary centres for genome research — two at the University of Washington in Seattle and one at Yale University in Connecticut. The new centres have grants totalling \$45 million over five years.

The University of Washington centres

will develop new devices for studying small populations of cells and will study how individual people's genomes differ, with the long-term goal of reducing the cost of sequencing a human genome from hundreds of millions of dollars to simply hundreds. The Yale centre will develop DNA microarrays to investigate where key regulatory proteins bind to the genome.

The awards are the first of about ten grants that the genome research institute plans to distribute over the next three years, as part of its plan to create a series of 'Centers of Excellence in Genomic Science'.

Australia names new research-agency head

Sydney Catherine Livingstone, former managing director of Australian hearing-implant company Cochlear, has been made chair of the country's largest research agency, the Commonwealth Scientific and Industrial Research Organisation (CSIRO).

The appointment of Livingstone, CSIRO's second consecutive head to come from a commercial background, is in harmony with the Australian government's plan to make the agency more dependent on industry funding. CSIRO's current



Catherine Livingstone: seeks industry money.

strategic plan is based on increasing its total funding, and in particular its external funding, which currently accounts for one-third of the agency's A\$870-million (US\$428-million) budget. CSIRO has shed around 900 staff since 1990, as government support for its programmes has declined.

♦ <http://www.csiro.au>

Push to commercialize superconductors heats up

San Diego The US Department of Energy has awarded \$57 million in grants for research to increase the use of high-temperature superconductors in commercial products.

The bulk of the projects in the four-year programme will deal with aspects of electrical power transmission, including the testing of new cable systems, generators and substations. New developments in magnetic resonance imaging and mineral-purification systems will also be studied. The firms and institutions receiving the grants will invest a further \$60 million in the projects.

Astronomy beefed up by MACHO database

Washington Astronomers and astrophysicists now have access to an online database of information about some 73 million stars in the Milky Way.

The database is a by-product of the Massive Compact Halo Objects (MACHO) project — an eight-year search for massive objects such as planets and brown dwarfs in the Milky Way conducted by astronomers from Australia and the United States. The site, which is free to access, includes images of the stars and records of their variability. The latter data are particularly useful as astronomers use them to measure distances.

"As these data are analysed by the world's scientific community they are certain to reveal some surprises," predicts Morris Aizenman of the US National Science Foundation, which helped to support the project.

♦ <http://www.macho.mcmaster.ca>

Erratum The News story of 30 August 2001 on mouse cDNA clones at Japan's Institute of Physical and Chemical Research (see *Nature* **412**, 848; 2001) incorrectly stated that Dnaform, a spin-off from the institute, is making the clones available to other researchers free of charge. Dnaform is in fact selling the clones at cost to customers conducting not-for-profit research.

Hard times for high tech

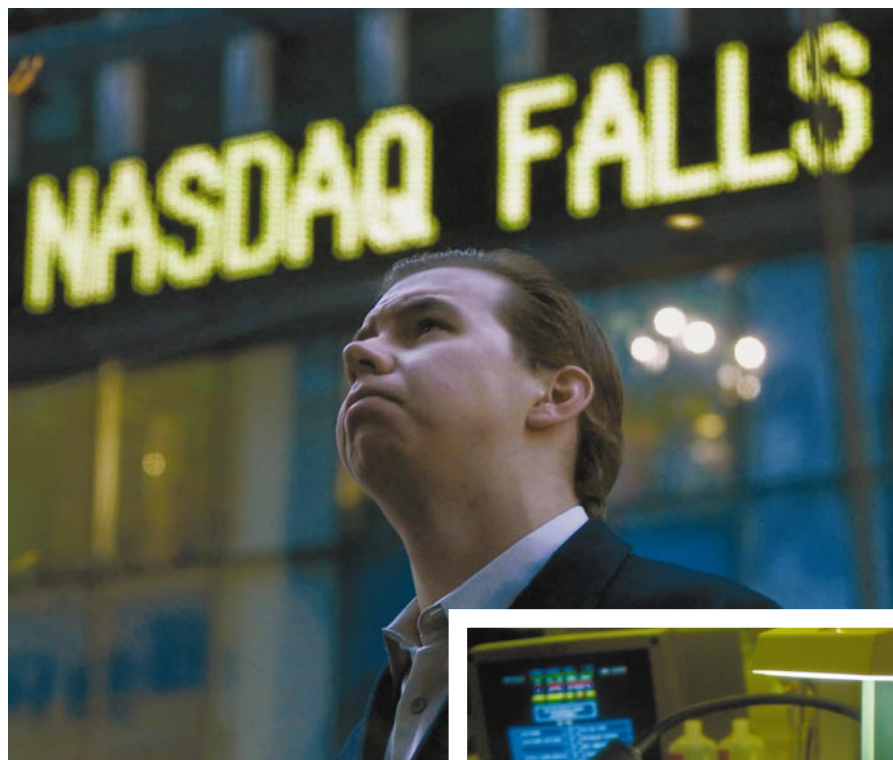
Computer and telecoms firms are losing money and laying off staff — yet say that R&D remains a priority. Can their scientists really remain immune to the economic downturn? Declan Butler and Jim Giles investigate.

Just two years ago, high-tech computing and telecommunications companies were on the crest of an economic wave. Mobile phones had hit the mass market and sales of computing equipment were buoyant, driven in large part by rapidly increasing Internet use. Investors rushed to embrace technology and dotcoms, and the Nasdaq high-tech stock-market soared to unprecedented levels.

In 2001, the bubble has burst in spectacular fashion. Nasdaq now lies in the doldrums, and high-tech firms' optimistic projections look seriously misplaced. Demand for mobile phones is down, and over-investment in the infrastructure of the Internet — optical-fibre networks and the equipment needed to run them — has left telecoms companies with a glut of unused technology. Computing sales are stagnant, and the chip market has become much more competitive.

Many analysts believe the global economy is on the brink of a full-blown recession. Factor in the economic fallout from the terrorist outrages in New York and Washington, and the outlook appears extremely gloomy. The worst-affected high-tech companies employ thousands of researchers. So what can the scientist at the bench expect to happen to his or her research budget?

In those companies that have made the worst misjudgements, researchers are deeply worried about the future. The British electronics company Marconi has a proud research tradition, tracing its roots back to the Wireless Telegraph and Signal Company, founded in 1897 by radio communications pioneer Guglielmo Marconi. But its decision to concentrate on telecoms has proved



BETH A. KEISER/AP

Cause for concern: will the collapse of high-tech stocks affect R&D in corporate labs?

disastrous, and there are fears for the company's survival. Last month, Marconi shelved plans for a new research facility and put plans for another under review. The firm is preparing to shed 10,000 staff. Between 35,000 and 40,000 jobs will be cut this year at Lucent Technologies, which runs the famous Bell Laboratories in Murray Hill, New Jersey. Bell Labs should escape the worst cuts, but morale is low (see *Nature* **412**, 578–579; 2001).

Most high-tech companies, however, are facing reduced revenues rather than outright ruin. In these, senior executives are keen to protect their R&D spending. Belts are being tightened, but research expenditure will not be the first to be cut. Some companies are even planning to increase investment in R&D, as the best way to ensure their continued competitiveness. Demand for the firms' current products may be static, but they know that they must be ready with the next generation when the economic gloom lifts.

Seizing the opportunity

"We have learnt that this is precisely the time that you want to be there with new products," says David Tennenhouse, director of research with the chip manufacturer Intel. He argues that a recession actually provides one of the best opportunities to pull ahead of the competition. With demand for the Pentium chips that sit inside the majority of personal computers having levelled off, Intel is now upping its R&D spending in order to focus on processors for mobile and wearable devices. On the software side, Bill Gates has



WILL & DENI MCINTYRE/SPL

been similarly bullish about R&D, saying that Microsoft's US\$3.8-billion investment last year will increase to \$5 billion this year.

This commitment to R&D is echoed at other leading high-tech firms. "The percentage of revenues we spend on R&D is going up," says Dennis Roberson, chief technology officer of Motorola, a leading player in mobile telecommunications and microchips. "I say this with a wry smile," he adds, "because revenues are actually falling faster than we are putting the percentage up. But R&D will not decrease by very much."

But these statements of support for R&D do not necessarily mean that researchers will be immune to the problems of the high-tech sector. Some analysts say that firms' commitment may waver if the downturn becomes a prolonged recession. Others suggest that the safety of research is relative and depends on the size and type of firm. Commercial R&D labs may also see more of their basic research farmed out to universities — accelerating a

trend that has emerged in recent years.

Even if the official R&D spending figures do remain stable, they may hide the full picture. R&D is notoriously difficult to define and the figures can be subject to creative accounting. Michael Blogg, an information-technology analyst at the London branch of investment bankers ING Barings, says that companies need to appear to be maintaining their R&D spending in order to satisfy investors. During previous recessions in the 1980s and 1990s, some companies reclassified other technical staff as R&D workers to hide cuts, experts say.

Threats to R&D are not spread out evenly across all companies. "Typically in a downturn, small firms are hurt more than large firms," says Sam Kortum, an economist at Boston University who specializes in technology and innovation. Companies with a wide range of products, such as IBM, can shield their R&D efforts from reduced revenues in one area. But smaller firms specializing in the worst-hit areas, such as telecoms networks and the electronics needed to run them, have nowhere to hide. Analysts warn that small companies trying to bring their first product to the market may go under.

Such companies survive on venture capital, and David Marino-Nachison, a writer for investors' website The Motley Fool, says that the current climate may force fund managers to shift their focus. "They'll be looking for companies that are further along the development cycle, or stand a chance of making a profit sooner," he says.

But small companies should fare better than in previous recessions, as the venture-capital industry now has more funding and better expertise. "The problem in the 1980s and 1990s was that the institutions were only just beginning to emerge," says Margaret Sharp, formerly a science-policy researcher and now a Liberal Democrat peer in Britain's House of Lords.

At larger companies, the downturn may accelerate recent trends in commercial R&D, which have seen firms demand that ideas are pushed much more quickly from research into marketable products. "There is a compression of timescales," says Luke Georghiou, director of PREST, the Policy Research in Engineering, Science and Technology group at the University of Manchester, who has just completed a study of how six of the largest high-tech companies manage R&D.

This means that the biggest pressures will be felt by staff working in product development — which accounts for more than 90% of high-tech firms' R&D. But another trend is affecting the bench scientists responsible for the 'R' part of this effort: a tendency to farm out basic research to academic groups or government research labs, rather than doing the work in-house.

Pooling resources

According to the latest edition of the Organisation for Economic Co-operation and Development's biannual report on science, technology and industry, published last month, industry now funds around 6% of all research in government and higher-education establishments, up 50% from the 1998 figure. "If the research really is basic and far from a product, we will seek expertise from universities or government labs," confirms Motorola's Roberson.

Lower revenues may also persuade firms to band together to fund basic research, says Paul Peercy, dean of the College of Engineering at the University of Wisconsin-Madison: "Firms are pooling resources and asking universities to look out for research that goes eight to ten years into the future." Peercy cites the example of the Semiconductor Research Corporation. Based in Durham, North Carolina, the corporation channels funds from chip makers, including Intel, Lucent and Motorola, into basic-research

programmes in US universities. The organization says it will distribute around \$30 million this year.

Researchers left behind in the corporate labs have mixed views about the impact of these developments. "There have been big changes," says one researcher at Agere Systems, formerly the microelectronics division of Lucent. "Managers are not so patient, they want to see a product in the short term." But those at companies that have suffered less than Lucent and Agere are confident that in-house fundamental research will still receive support. "We have been asked to think about spin-offs," says a scientist doing basic research with Btexact Technologies, the research arm of British Telecommunications. "But we are still able to tackle the same issues. Our work could have a big pay-off, so all telecommunications companies will have to have at least some people looking at this stuff."

Ultimately, the length of the downturn may determine how corporate R&D will fare. Financial forecasting, always an imprecise science, will be made doubly difficult by the global instability stemming from the terrible events of 11 September. Some executives, Motorola's Roberson among them, admit their plans could change if the climate does not improve within the next few years. Others, such as those at Intel, say they will protect R&D come what may. Commercial realities will inevitably intervene at some point, however, and few believe that R&D budgets could be maintained at current levels through a deep and prolonged recession.

But with investors keeping an eye on R&D spending, and executives keen to keep the product pipeline flowing, R&D budgets look safe for now. As Tom Theis, director of physical science research at IBM, says: "You're either investing for the future, or you're not."

Declan Butler is *Nature's* European correspondent; Jim Giles is *Nature's* assistant News and Features editor.



Different outlooks: R&D at Marconi (far right) has been hit hard, but budgets look safer elsewhere.



Biologists join the dots

Tiny specks of semiconductor can make biological molecules and cellular components glow in a kaleidoscope of colours. These 'quantum dots' may soon be a standard biological tool, says Erica Klarreich.

Take some crystals of cadmium selenide, measuring a just few billionths of a metre across, add an ordinary microscope, and what have you got? According to a small but growing band of enthusiasts, this simple recipe could bring about a quantum leap in cellular imaging and biosensing.

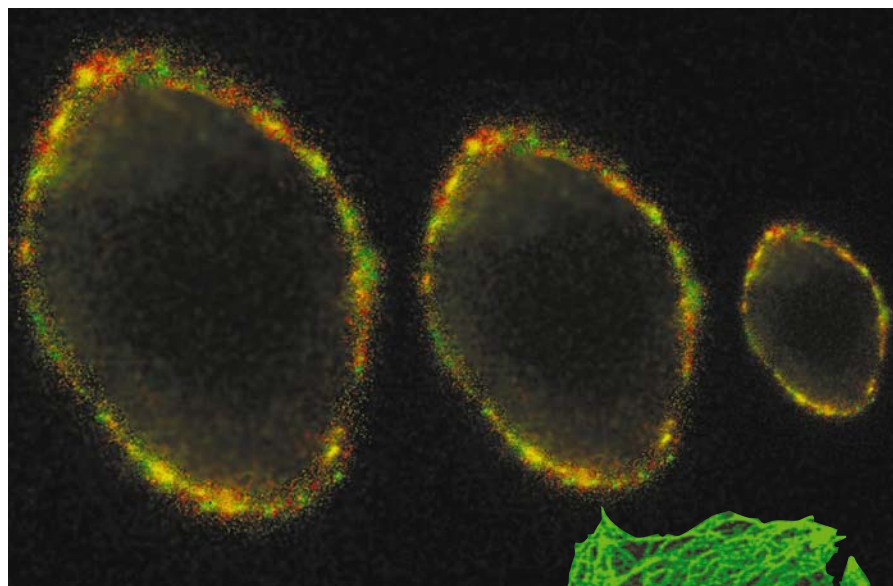
Nanocrystals of semiconducting materials, otherwise known as quantum dots, have fascinated physicists, chemists and electronic engineers since the 1970s. The dots can confine electrons in a tiny three-dimensional space. Excite these electrons with a beam of light and they re-emit light of a precisely defined wavelength as they drop back to lower energy levels. The dots' small size gives them unusual properties that have sent researchers scurrying to investigate potential applications in electronic and optical devices — and even in a future generation of quantum computers.

Yet quantum dots are now poised to find their first practical use in the biology lab — potentially doing for cell biology and other disciplines what Technicolor did for Hollywood. When excited by a pulse of light, quantum dots can be made to glow in a variety of distinct colours so intense that a relatively unsophisticated optical microscope can spot a single dot. If attached to molecules that home in on specific biological structures, quantum dots could shed light — in a rainbow assortment of hues — on the inner workings of cells.

The promise does not end there — quantum dots could be used to screen candidate drugs or to provide a rapid read-out of the proteins present in a particular cell or tissue. "They've got great potential," says Carolyn Larabell, a cell biologist at the University of California, San Francisco, who has been studying how cells take up the dots. "We're just at the start of finding out what secrets they can reveal."

Although researchers are still perfecting the dots' surface chemistry, quantum dots could find their way into cell biologists' toolboxes within a year or two. Two companies — Quantum Dot Corporation in Hayward, California, and BioCrystal in Westerville, Ohio — are already preparing to market them for biological applications.

One of the main attractions of quantum



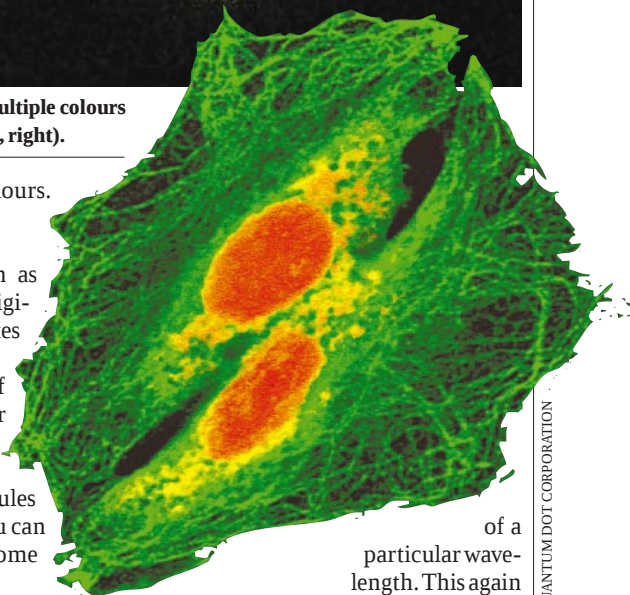
Going dotty: quantum dots can stain in multiple colours (above) and label cell microtubules (green, right).

dots is the wide range of available colours. Tracking biological molecules using fluorescence is far from a new idea, but conventional dyes — such as green and red fluorescent proteins, originally isolated from marine invertebrates — suffer from a serious drawback: they each emit light over a range of wavelengths, which means that their spectra may overlap. This makes it difficult to use more than one dye at a time to tag different biological molecules simultaneously. "If you are careful, you can get two colours," says Larabell. "Some people say they can see three, but that's very tricky."

Whatever hue want

Quantum dots, on the other hand, emit light at a variety of precise wavelengths, the exact colour depending on a dot's size (see 'Pick a colour', opposite). Shuming Nie, a quantum-dot chemist at Indiana University in Bloomington, for example, has produced dots that fluoresce in more than a dozen different colours. In principle, experts agree, it should be possible to make even more.

Another disadvantage of conventional dyes is that they can only be activated by light



of a particular wavelength. This again limits the number of dyes that can be deployed in unison — if you simultaneously blast a cell with a series of lasers tuned to different wavelengths, it is likely to be damaged by the beams' cumulative energy. But a quantum dot can be excited by any wavelength shorter than the one it emits, which means that a series of different-coloured dots can be activated using a single laser. "Since you can use one laser for multiple probes, it's potentially less damaging," says Larabell.

Quantum dots are also extremely stable, and can go through repeated cycles of excitation and fluorescence for hours at a time. Conventional dyes last only for

minutes, and fade rapidly if they are continually excited. The longevity of quantum dots offers the possibility of making movies of long-term interactions of biological molecules in a living cell, each biomolecule being tagged with a different colour.

This could transform the way in which cell biologists design their experiments, says Paul Alivisatos, a chemist at the University of California, Berkeley, and one of the founders of Quantum Dot Corporation. "Biologists have always had to figure out what they could do in the first 5 or 10 minutes before the dye ran out," he says. "Now they can look for much longer."

Nanocrystals of cadmium selenide can be made by mixing dimethyl cadmium and selenium in a high-temperature surfactant solution. The surfactant molecules prevent the growing crystals from merging. Once the dots have reached the desired size — sometimes containing just a few dozen atoms — the solution is cooled, making the surfactant molecules coat their surfaces more effectively and so halting their growth.

Rather than using this technique, those interested in the fundamental physics and chemistry of quantum dots, as well as engineers trying to integrate them into electronic devices, instead generally produce them as part of a solid structure, in which the nanocrystals are sandwiched between layers of a different material. Assembling these structures into electronic devices poses a tough challenge, which may explain why biological applications currently lead the way. "It's very difficult to build transistors and complex circuits out of quantum dots," says Shimon Weiss of the University of California, Los Angeles, and a founder of Quantum Dot Corporation. "For biology you don't have to link them together."

For the dots to be useful for biological imaging, however, they have to be coated with substances that will make them home in on specific biomolecules. In 1998, two research groups set the ball rolling by revealing how to link quantum dots to molecules that guide them to important cellular structures.

One team, led by Alivisatos and Weiss, figured out how to attach red dots to a protein called biotin which, after forming a chain of interacting molecules, then labelled filaments of actin protein in mouse fibroblast cells. The group also linked green dots to urea and acetate groups, which carried the dots into cell nuclei, directly 'staining' them¹. In independent work, a team led by Nie linked dots to antibodies and then used them to label specific target proteins in human cancer cells².

A wide range of molecules could serve as useful guides for quantum dots. These include nucleic acids; lipids that attach to cell membranes; proteins with affinities for certain receptor proteins or sugars; and drugs,



Paul Alivisatos believes the dots will transform cell biology.

many of which target particular cellular structures.

At Vanderbilt University in Nashville, Tennessee, a team of chemists and neuroscientists is working out how to use quantum dots to aid studies of neurotransmitters, the chemical messengers that co-ordinate the firing of our nerve cells. Using dots linked to the neurotransmitter serotonin, the researchers are studying the behaviour of the transport protein that pulls the neurotransmitter back into a cell after it has transmitted a signal across the gap between adjacent nerve cells.

Similar dots might also prove to be an efficient tool for screening candidate drugs, says chemist Sandra Rosenthal, a member of the team. Often, a successful drug must be able to bind to several different molecular targets to achieve the desired effect, and steer clear of other targets to avoid side-effects. Attaching different-coloured quantum dots to the various targets could make testing a simple matter, Rosenthal says. "A good hit might be a drug that displaces, say, blue, aqua and green nanocrystals where you want it to attach, but doesn't displace red, yellow and orange ones at proteins that indicate side-effects."

Quantum dots may even find a niche in medical imaging — although this will

require dots that glow in the infrared spectrum. "Tissue is more or less transparent to infrared light," explains Marcel Bruchez, a senior scientist with Quantum Dot Corporation. "It can penetrate to centimetres, where with normal light it's millimetres at the most."

Plugging the gaps

This means looking beyond the current material of choice, cadmium selenide, which only emits light in the visible spectrum, and embracing other semiconductors, such as indium arsenide. Less is known about how to coat crystals of indium arsenide with molecules that will make them home in on particular biomolecules, warns Alan Waggoner, a specialist in fluorescence imaging at Carnegie Mellon University in Pittsburgh. But he is now collaborating with Quantum Dot Corporation to try to solve the problem. "It's just chemistry — you keep plugging away and sooner or later you get an answer," he says.

If the dots are to be used in humans, Waggoner adds, researchers will also have to figure out how to get our bodies to eliminate them. "You would want these things to be excreted eventually," he says.

Nie, meanwhile, is packing different combinations of quantum dots of various colours into latex beads a millionth of a metre in diameter, in carefully controlled ratios. His goal is to create a system of 'spectral fingerprints' that will allow the detection of specific genetic sequences or proteins in high-throughput biological assays. Using



Pick a colour

Quantum dots are nanometre-scale semiconductor crystals. When semiconductors are hit by a beam of light, some of their electrons are excited into higher energy states. As the electrons fall back to the ground state, they emit photons of light in a colour that is characteristic of the material.

When the electrons in a semiconductor are excited, they prefer to dwell at a certain fixed distance from the positive charges they leave behind, called the exciton radius. For the semiconductors used to make fluorescent quantum dots, this radius is around 5–10 nanometres.

If the entire crystal's size is less than the exciton radius, however, an effect called quantum confinement comes into play, and shifts the colour of the emitted light towards shorter wavelengths. For crystals of this size, more energy than normal is required to force electrons out of the ground state. So when the electrons return to the ground state, the photons they release have more energy — and therefore a shorter wavelength — than normal.

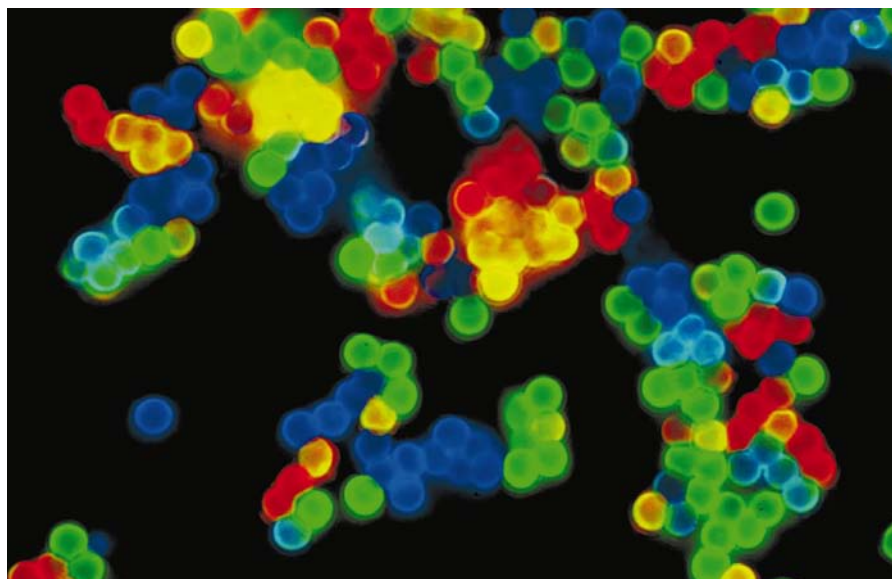
There is a simple linear relationship between crystal size and colour: the smaller the size, the shorter the wavelength. For cadmium selenide — one of the most popular materials for making fluorescent quantum dots — varying the size of the crystals over a range of 2–6 nanometres covers the entire visible spectrum.

dots emitting three distinct colours at ten different intensity levels, Nie already has created hundreds of beads with distinct fingerprints³. With six or seven colours, it should be feasible to extend the number of spectral fingerprints to 10,000 or more.

These beads could then be tagged with short genetic sequences to make them bind to specific DNA sequences or to the messenger RNA (mRNA) produced by active genes. The beads could also be tagged with antibodies so that they recognize specific proteins. By mixing a liquidized biological sample with a solution of beads tagged in this way, the beads could be used to provide a rapid read-out of the genes that are active, or the proteins that are present, in the sample.

The idea would be to label all of the mRNA or protein in a sample with a conventional dye, or a quantum dot of a particular colour. After the sample is mixed with the quantum-dot-labelled beads, the beads would be fed through a flow cytometer, a device that funnels them one-by-one across a small opening. Here, a laser-based system would read the beads' spectral fingerprints and identify the beads that had bound to a target molecule.

Currently, high-throughput analyses of gene expression are conducted using 'chips' or microarrays, glass slides on which hundreds or thousands of DNA sequences are immobilized. Biological samples in which all the mRNA sequences have been



Dot detectives: latex beads with 'spectral fingerprints' of dots could be used to detect gene expression.

fluorescently tagged are washed over the chip, and lasers are used to determine where on the chip the mRNAs bind — showing which genes were active in the sample. Nie argues that his approach should provide results more quickly. "DNA chips require about 24 hours," he says. "The beads would just take 10–15 minutes."

Flexible friends

Quantum-dot beads would also be more flexible. In the microarray approach, adding new genes to the analysis means producing a new chip. But analyses in solution have no such limitations. "You just dump in the beads, and you can add new ones or remove old ones as you like," says Alivisatos. "It just might make the difference between a few people being able to do these assays at great expense, and people being able to do them very widely."

In the long run, says Nie, the main benefit could be in the emerging field of proteomics, as it is proving difficult to design 'protein chips' to screen for the presence of proteins in the way that DNA microarrays screen for gene activity. Quantum-dot beads tagged with antibodies might provide the answer, Nie suggests. Unlike flat chips, beads in solution could allow antibodies to maintain their normal three-dimensional structure, which is crucial to their functioning.

But it is in the field of cell biology that quantum dots may have their biggest impact. For biologists who want to monitor the interactions between biological molecules or cellular components over long periods of time, says Alivisatos, quantum dots could "change their whole mode of operation".

For instance, it may become possible to determine in living cells whether two proteins are linked together into a complex. When proteins tagged with dots are phys-

ically linked in this way, the dots are close enough together that light emitted from one dot can activate the other, making it glow more brightly than normal. "This could give us information about the dynamic interaction of two proteins," says Randy Blakely, a neuroscientist at Vanderbilt University who works with Rosenthal. "The optical resolution of a normal fluorescent microscope isn't enough to tell whether molecules are physically interacting with each other."

Some technical problems remain to be ironed out, however, mostly relating to the dots' surface chemistry. "We need to protect the quantum dots so that they are not oxidized, are very stable, can withstand salt concentrations in cells, and don't bind to what you don't want them to," says Weiss. The very versatility of quantum dots — the fact that their cellular target can be changed simply by changing their surface chemistry — is both a blessing and a curse, Bruchez adds, as it means that the surface must be tailored for each biological application. "We're looking for a robust surface that is widely applicable," he says.

Although neither BioCrystal nor Quantum Dot Corporation has figured out yet how to mass-produce the dots cheaply, both have started making their dots available to small groups of researchers, and plan to expand their distribution in the coming year.

Once the dots become more widely available, Alivisatos says, it might take biologists a while to explore their potential. But Larabell is confident that this won't take long. "When the methods are worked out, they'll be used instantly," she predicts. ■

Erica Klarreich is an intern with Nature.

1. Bruchez, M. Jr, Moronne, M., Gin, P., Weiss, S. & Alivisatos, A. P. *Science* **281**, 2013–2016 (1998).
2. Chan, W. & Nie, S. *Science* **281**, 2016–2018 (1998).
3. Han, M., Gao, X., Su, J. & Nie, S. *Nature Biotechnol.* **19**, 631–635 (2001).



Screen saver: Shuming Nie says the dots could solve the problem of screening for many proteins.

Rules on originality need to be clearly set out

Sir — Japan's Institute of Physical and Chemical Research, RIKEN, has now published its investigation into US accusations of espionage against Takashi Okamoto, head of the institute's laboratory for neurodegeneration signalling (see *Nature* **411**, 225; 2001 and *Nature* **411**, 991; 2001).

RIKEN's report, whose main author is chief executive Shun-ichi Kobayashi, concludes that there is no evidence that the institute intentionally planned to bring in the materials from the United States. But Okamoto has remained silent on the issue, and has now abruptly left RIKEN.

This affair has shocked Japanese scientists and highlights the differences in definitions of 'originality' between Japan and the United States. The United States encourages excellent scientists from all over the world to work there on short-term contracts, and is understandably keen to maintain its competitive edge. But to avoid similar misunderstandings in future, the US government should issue clear guidelines for foreign researchers about originality and authorization for removal of research materials.

Tadasu Nagaoka, Hisatsugu Miyakoshi

Kouseiren Takaoka Hospital, Eiraku 5-10, Takaoka, Toyama, 933-8555, Japan

Related problems

Sir — In his Words essay "Yes, but what's it for?" (*Nature* **412**, 771; 2001), Steve Blinkhorn reminds us that the current state of language can make it difficult to discuss evolution in an accurate way. We must be constantly alert to invocation of teleological explanations when describing evolution.

I believe that another (related) problem lies in the inherent ambiguity of the term 'related', used to depict similarity between biological objects such as nucleic acid or protein sequences. It is common to see very similar sequences described as "closely related", less similar ones as "distantly related" and dissimilar ones as "unrelated". Of course, the truth is that related (similar) sequences are not always related in terms of shared evolutionary origin (homology), and vice versa.

This ambiguity in meaning of the word 'related' differs from that which some believe exists for 'homology'. It is generally agreed that the only useful definition of homology between biological objects is "having a common evolutionary origin" (G. R. Reece *et al. Cell* **50**, 667; 1987). But when using the term 'related' in a biological context, it would be helpful to

readers if authors could be explicit about which meaning is appropriate.

Alex C. W. May

Division of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

Give us time to put reforms into action

Sir — In your News feature "Forza Scienza!" (*Nature* **412**, 264–265; 2001), you discuss the scientific management of Italy's Higher Institute of Health (ISS) in

Rome since its reform. Yet this reform will not be operative until the end of the year, when the rules and regulations set down in March will be approved. Therefore it would seem wiser and more appropriate to leave us some time to implement the reform before quoting statements such as "The situation [at ISS] is worse than ever".

The whole staff, with me at the front line, is strongly committed to exploit all the autonomy and flexibility provided by the new rules to allow us to strengthen our research and output in terms of public health. This is our only mission.

Enrico Garaci

ISS, Viale Regina Elena 299, 00161 Rome, Italy

In the stem-cell debate, new concepts need new words

Confusion arises when the existing vocabulary we use is inadequate to describe new methodologies.

Sir — Your Opinion article "The meaning of life" (*Nature* **412**, 255; 2001) implied that the suggestion made by Advanced Cell Technology's ethics advisory board, to substitute the term 'ovumsum' for 'embryo' in the case of eggs activated following nuclear transfer, is an attempt to sidestep the moral issue associated with deriving stem cells from human embryos. This implication is not correct.

I am the board member who derived the term 'ovasome' (not 'ovumsum') for reasons of accuracy of language, not out of a desire to sidestep controversial issues. Much of the current confusion in the US Congress stems from describing new methodologies with existing language.

The heart of the matter rests with the wondrous properties of eggs. Because, as you point out in your article, neither eggs nor sperm are dead before fertilization, a compelling argument has been made that fertilization is not the beginning of life, but a continuum of life in a new form. The real issue is, therefore, the perceived threat to the sanctity of the union of egg and sperm to create a new individual.

Unfortunately, the union of human eggs and sperm fails far more often than it succeeds in producing a new being. It is this high failure rate, and the inability to prevent it, that has led to the horrifying stores of frozen human embryos worldwide. Rather than asking if these frozen embryos can be used to derive stem cells, we should be asking why so many were created. And should more be created for the sole purpose of deriving stem cells, as was done recently in a Virginia clinic?

The answer is no. Other technologies

can be developed to produce stem cells with greater medical advantage. Stem-cell therapy holds the promise of replacing defective cells in adult organs, and the optimal way to attempt this medically is with cells from the affected individual. Eggs are capable, by an as-yet unknown mechanism, of remodelling the nucleus of an adult cell into the much larger, more open format of a 'pronucleus', the first nucleus formed by an activated egg. This yields a new cell with vastly expanded potential to develop into a variety of cell types: this is the power to be harnessed.

An alternative approach is to activate the egg with its own genetic material intact (parthenogenesis). Stem cells derived from parthenotes will have half as many problems of tissue rejection as stem cells derived from eggs fertilized by sperm, which will have the same tissue compatibility problems as transplanted organs.

The early stages of developing these procedures may be similar to those used to clone individuals, but as time goes on the efficiency of creating genetically matched stem cells from activated eggs will increase, and the potential for development to an offspring will decrease. It follows that if creating an offspring is not the goal, the term 'embryo' is not accurate. 'Cloning' is more accurate, but is broadly used to describe a variety of experimental procedures. Hence my suggestion of 'ovasomagenesis' to describe the process and 'ovasome' to describe the product of activating an egg to become a somatic cell.

Ann A. Kiessling

Harvard Medical School, Blackfan Circle, Boston, Massachusetts 02115, USA

Whither evo–devo?

Assessing the interplay between evolutionary and developmental processes.

From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design

by Sean B. Carroll, Jennifer K. Grenier & Scott D. Weatherbee

Blackwell: 2001. 214 pp. £29.95, \$44.95

Axel Meyer

Did you ever wonder how snakes lost their necks and legs? Are you curious about why birds have feathers on their wings and scales on their legs? You will find the answers to these and many other intriguing questions in this excellent book, in which Sean Carroll, Jennifer Grenier and Scott Weatherbee describe the state of 'evo–devo' in an admirably lucid style.

The past ten years have seen a renaissance of interest in the connections between developmental and evolutionary biology and the rebirth of a biological discipline colloquially called 'evo–devo' (or 'devo–evo', depending on where you're coming from). Previous incarnations of this field by earlier generations of biologists lacked molecular genetics — now this discipline enjoys top billing. The most surprising result of this young and extremely fast-moving field (in which several new journals have been started and many universities have established professorships) has been the realization that a truly amazing developmental and genetic commonality underlies all the diversity of animal types.

With hindsight, perhaps, this conservation, which includes the almost universal sharing of genetic pathways specifying developmental processes, should not have been so surprising; it is almost trivial to state that the genome of a species contains a record of its entire evolutionary history and those of all its ancestors. But although this might sound like a truism, its implications are manifold and deep, and have led to a greater understanding of how animals develop, and how the commonality in their genomes and developmental mechanisms connect the very different sets of characteristics, or phenotypes, of distantly related phyla.

From DNA to Diversity is written for a general audience, including undergraduates, with an interest in developmental and evolutionary biology, and it is a joy to read. Using striking examples, the authors brilliantly summarize the current state of thinking on the interconnectedness between developmental genetics and evolutionary diversification.

For example, they describe how ideas about the evolution of eyes have been transformed by the new perspective. Eyes were thought to have evolved independently

dozens of times, not only because some phyla that have eyes are more closely related to phyla without eyes than to other phyla that have them, but also because the structure of animal eyes is so varied. An astonishing example of evolutionary conservatism, however, involves the function of the *eyeless* gene in flies and *Pax6*, the corresponding gene in mice. Despite more than 600 million years of separate evolution of flies and mice, the introduction of the mouse gene into flies can induce new eye tissue — not of the camera-like eyes of mammals, but of the insect compound eye!

This conservation raises several important questions. How can such different structures, all serving to detect light, be made through the control exerted by essentially the same gene in different phyla? *Pax6/eyeless* sits near the top of a regulatory chain of genes involved in the development of all types of eyes, even those differing hugely in morphology. *Pax6* and its functions have been conserved throughout the Metazoa, and it was probably present in the genome of the ancestor of all bilaterians. Given that some phyla have never developed eyes, why were these genes not lost during evolution? The selective forces pushing genomes to be small and lean do not seem to be too strong. It is likely that this cascade of gene interactions was co-opted again and again into making eyes, having served other developmental functions in the (sometimes extremely long) evolutionary meantime. The repeated *de novo* assembly of this network, gene interaction by gene interaction, is implausible indeed.

Although not touched upon by Carroll *et al.*, this raises other important questions about homology, one of the central themes in biology. Do structures have to be simply either homologous or non-homologous, or can they be partly homologous, depending on whether or not their ancestors possessed these structures and/or whether a homologous or non-homologous developmental pathway is used to make them? It is becoming clear that there are different kinds of homologous relationships and that homology can exist at different levels of the biological hierarchy, from genes to structures. The distinctions between homology (similarity of structures due to evolutionary descent) and homoplasy (similarity due to similarity in function and not evolutionary descent) are becoming increasingly blurred.

Classic issues about convergence in phenotypes can now be studied at the genomic and developmental level to investigate whether evolution invented the wheel



Of scales, feathers and skin: the evo–devo toolkit that makes a snake a snake, and not a bird.

twice or found two different ways of making a similar wheel. It seems likely that, more often than not, the answer will be that old gene cascades are re-recruited, never having really been lost, to provide a similar phenotypic solution to a comparable ecological challenge. Evolution is a tinkerer that uses the tools at its disposal rather than repeatedly starting from scratch.

Morphology is the product of development, which is controlled by genetic regulatory programs — genetic toolkits, as Carroll *et al.* term them. Two major groups of gene products underlie these genetic programs — transcription factors, which regulate the expression of other genes, and signalling

JEFF FOOTY/BBC WILD

proteins, which mediate short- and long-range interactions between cells. The best-known examples of developmentally important transcription factors are the family of Hox proteins, which are involved in, among many other things, specifying segmental identity along the antero-posterior axis. The interactions between the Hox genes themselves are extremely complex, involving at least six classes of regulatory elements, in addition to autoregulatory feedback mechanisms.

This tight regulatory coupling is probably one of the main reasons for the extreme evolutionary conservation of the architecture of the tightly linked Hox gene cluster. Hox genes lay down a ground plan in animals that translates amazingly well into major features

— segments or modules — of their morphology. The universality of the Hox gene clusters has led to the — now often questioned — suggestion that the presence of a linked set of Hox genes is a defining characteristic of all animals, the so-called 'zootype'.

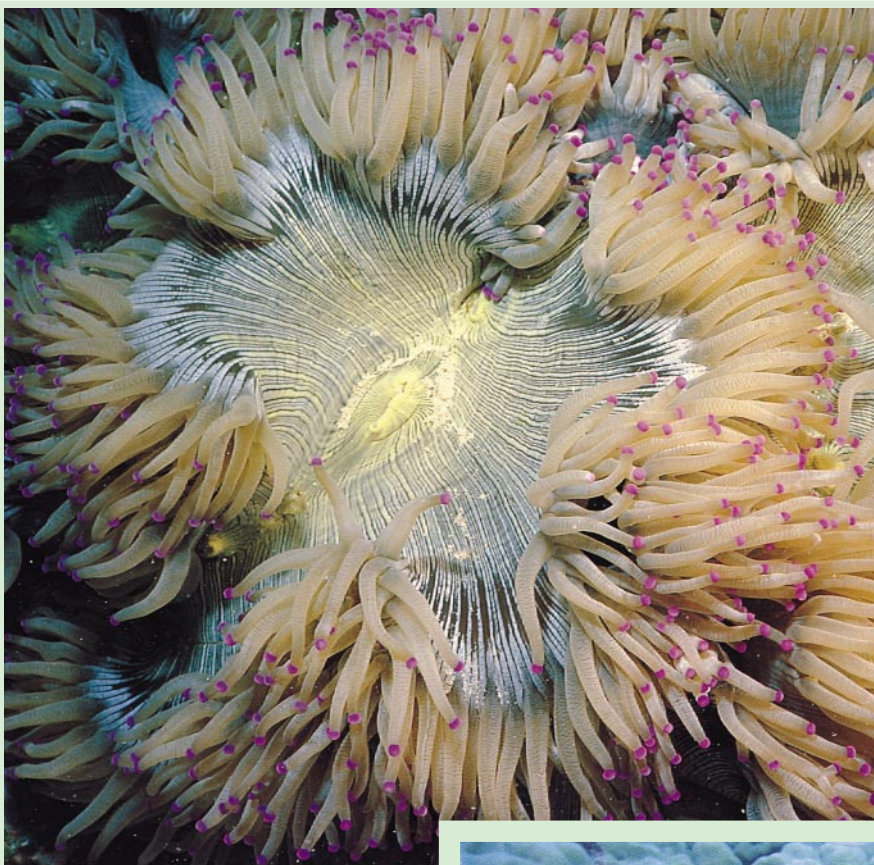
Evolution can select only for what is developmentally possible. The evo-devo field has matured, and now that we know conservation abounds, attention must focus on the big question — how does evolution make new and different things? Given the overwhelming similarity and stasis at many levels, ranging from genes and genomes to the interactions of genes in gene networks, how do differences between species arise? How can phenotypic differences be explained, and

what, if any, are the rules of change? Carroll *et al.* argue that the answer probably lies largely in changes in the regulation of gene expression. I think that other kinds of molecular mechanism, such as alternative splicing, ribosomal RNA editing and gene duplication, will also be found to have major roles in explaining phenotypic diversification.

With regard to the influence of gene regulatory mechanisms in evolution, the Hox genes are prime examples of how selective and differential regulation of gene expression can confer distinct identities on body segments that were originally serially homologous. Such 'individuation' of segments, as Carroll *et al.* call it, results from the partial uncoupling of the underlying developmental program of each segment from the gene networks controlling the development of segments with other identities.

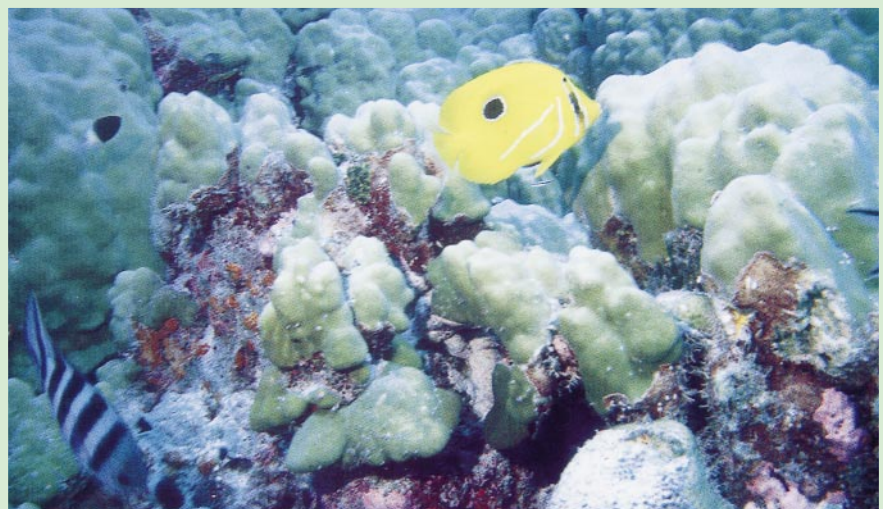
Carroll *et al.* describe how an increasing amount of work on the modular nature of the often extensive and complex regulatory regions of Hox and other developmental genes shows that it is neither accurate nor sufficient to try to explain the functions of a given evolutionary toolkit only in terms of the proteins it includes. The function of a protein always depends on the spatio-temporal context of its expression, which can be altered by changes in one or more of its gene's regulatory modules. A protein function can thus be dissociated from its original spatial and temporal pattern of expression, enabling the evolution of new gene and protein interactions, and thereby the evolution of phenotypic novelty or the individuation of particular segments and developmental modules.

Individuation of a module, such as the conversion of the gill arches of fish into functionally new structures such as jaws, or the conversion of arthropod segments used in locomotion into segments used for feeding structures, antennae or genital structures, is the stuff of evolutionary novelty. Modularity



The shrinking world of corals

Australia's Great Barrier Reef hosts the elegance coral, *Catalaphyllia jardinei*, shown above. The picture comes from *Corals of the World* by J. E. N. Veron (3 vols; Australian Institute of Marine Science, A\$265). In the Indo-Pacific, the Bennett's butterflyfish, *Chaetodon bennetti* (right), feeds primarily on coral polyps, and is among the species affected by the mass coral mortality — from *World Atlas of Coral Reefs* by Mark D. Spalding, Corinna Ravilious and Edmund P. Green (University of California Press, \$45, £29.95), which details the state of the world's coral reefs.



and repetition of parts coupled with individuation is the emerging principle of phenotypic organization, and is reminiscent of the organization of genes and the structure of genomes. Are there generalities, or even rules, of evolutionary change to be discovered here? Carroll *et al.* contend that these findings can explain why modularly organized animal phyla such as arthropods are among the most diverse in terms of numbers of species and morphological diversity.

'Hopeful monsters' — the hypothetical outcome of a major change in morphology at a single step — are fated to be evolutionarily stillborn. Macroevolutionary innovation must have its roots in microevolutionary variation in development at the population level, and this is going to be a major area of future research in the evo-devo field. All this and more is described in this eminently readable book, which is illustrated with beautiful colour figures. Any new graduate student in either developmental or evolutionary biology should read it. The book has the potential to inspire the next generation of biologists to use the developmental and genomic tools now available to address some of the most exciting questions in biology.

To paraphrase the Apple Powerbook commercials: "You so gonna want this book."

Axel Meyer is in the Department of Biology, University of Konstanz, 78457 Konstanz, Germany.

More on evo-devo

Cycles of Contingency: Developmental Systems and Evolution

edited by Susan Oyama, Paul E. Griffiths
& Russell D. Gray
MIT Press, \$50, £34.50

No easy answers

Human Trials: Scientists, Investors, and Patients in the Quest for a Cure

by Susan Quinn
Perseus Publishing, 2001. 295 pp. \$26, £18.99
Xavier Bosch

Until the early 1950s, the effectiveness of a clinical therapy was judged largely by doctors' opinions rather than by statistically rigorous observational evidence. The modern clinical trial has been essential for the movement to a type of medicine where treatment is expected to be based on firm evidence of benefit rather than opinion. Although far from perfect, present-day clinical trials follow a strict scientific methodology to find out whether a new treatment option is safe, effective and a better alternative to current treatments.



Weiner at work: his research into autoimmunity led to a clinical trial for treating multiple sclerosis.

In *Human Trials*, Susan Quinn gives us a readable and authoritatively written insight into what is at stake when a determined and talented scientist puts his ideas on the line in a clinical trial. She relates the efforts of Howard Weiner, Harvard University neurologist and scientist, to find a useful treatment for multiple sclerosis (MS). In this relapsing, and often eventually progressive, autoimmune disorder of the central nervous system, the myelin sheaths surrounding nerve-cell axons become the target of immunological attack.

Weiner's proposed treatment turns on an approach known as oral tolerance, a relatively 'natural' and logical therapy aimed at suppressing an immune response to a given antigen by previous administration of the identical or similar antigen by the oral route. The scientific rationale behind the idea is that, by stimulating the immune mechanisms in the gut-associated lymphoid tissue (GALT) of the small intestine, it is possible to suppress and interrupt the autoimmune disease process. Because the ingested antigen is naturally derived — it is a natural component of the body — and undergoes the normal digestive processes, the approach would seem to be remarkably safe. And it has the great advantage that the antigen can be given simply in the form of a pill.

Quinn describes how and why the oral-tolerance idea, discredited and regarded almost as homeopathy by some, was put into

practice to find out whether it works in multiple sclerosis and rheumatoid arthritis. After successful results in laboratory mice given an oral myelin preparation, and in a small group of MS patients, Weiner carries out a phase-I clinical trial to test safety and then decides to test the efficacy of his treatment on a large group of MS patients in a phase-III multicentre clinical trial throughout North America.

Weiner's enthusiasm and faith in the natural, oral approach is transmitted to his colleagues in the laboratory and the clinic and also to venture capitalists, who help him to found a biotech start-up company, which goes public. The optimism, strengthened after publication of the preliminary findings in prestigious journals with notable media impact, is also passed on to the patients, who face enormous prognostic uncertainty and have few therapeutic options. Quinn conducted extensive interviews with patients and scientists during the course of the trials, observed meetings at the company and read Weiner's personal diaries.

Although encouraged by the initial results of the MS trial, Weiner has reservations about the doses used and the type of antigen administered. There is also some concern that, because of time constraints, a phase-II trial had been skipped to enter phase III directly. Phase-II trials involve a larger group of patients than phase I and

researchers build on what they learned in phase I. Because more subjects are involved, investigators can discover less common side effects, and can continue to evaluate the safety of the new approach.

Quinn cleverly maintains a high level of suspense until the statisticians reveal the final results. The future of the company, and the credibility of the approach and of the whole idea is at stake. But, despite signs of considerable improvement in some patients, the news was bad. Overall, the treatments failed to exceed the placebo effect in both the multiple sclerosis and arthritis trials, with devastating effects for the company.

In the course of recounting this dramatic story, Quinn discusses such issues as the extremely sensitive nature of biotech shares to news, and the attempts to interpret statistics in a way that oversells results. As well as providing a useful insight into how science moves from the lab to the clinic via the stock market, she also covers the human aspect — the relationships between patients and doctors and between scientists and biotech executives. She explains the ethical concerns raised by some clinicians about the requirement that disease-modifying agents such as methotrexate, a cytotoxic drug used to treat autoimmune conditions, had to be withdrawn from patients in remission before they could start the trial treatment — with a 50:50 chance of receiving the placebo.

There is perhaps too much time spent on Weiner's personal and family history in the first part of the book, but *Human Trials* is clearly aimed at the general reader and

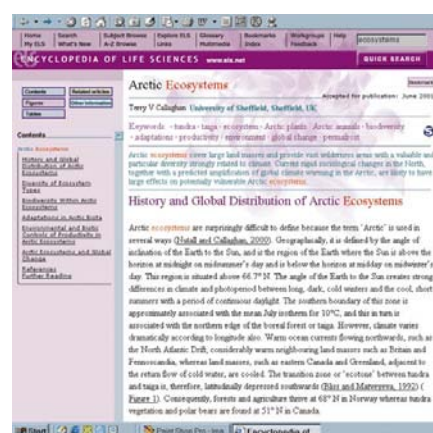
perhaps even the potential biotech investor. Clinicians interested in how clinical trials are carried out, and scientists involved in autoimmune-disease research, in particular in disorders where oral tolerance may have potential, should also enjoy reading it. ■

Xavier Bosch is in the Department of Internal Medicine, Hospital Clinic, University of Barcelona, Villarroel 170, 08036-Barcelona, Spain.

Is content king?

The Encyclopedia of Life Sciences
(www.els.net)
Nature Publishing Group: 2001. 20 vols.
See website for prices
Johanna McEntyre

The electronic publishing of encyclopaedias has had a unique recent history. The story of the rise, fall and possible redemption of *Encyclopaedia Britannica* in particular is documented in *Blown to Bits: How the New Economics of Information Transforms Strategy* by Philip Evans and Thomas S. Wurster (Boston Consulting Group, 1999). This household name was at the peak of its powers in 1990, with revenues of about \$650 million. But sales of printed encyclopaedias were then decimated by more than 80% by the arrival of the CD-ROM. The sheer size of the *Encyclopaedia Britannica* (too big to fit on a single CD-ROM) and the belief of *Britanni-*



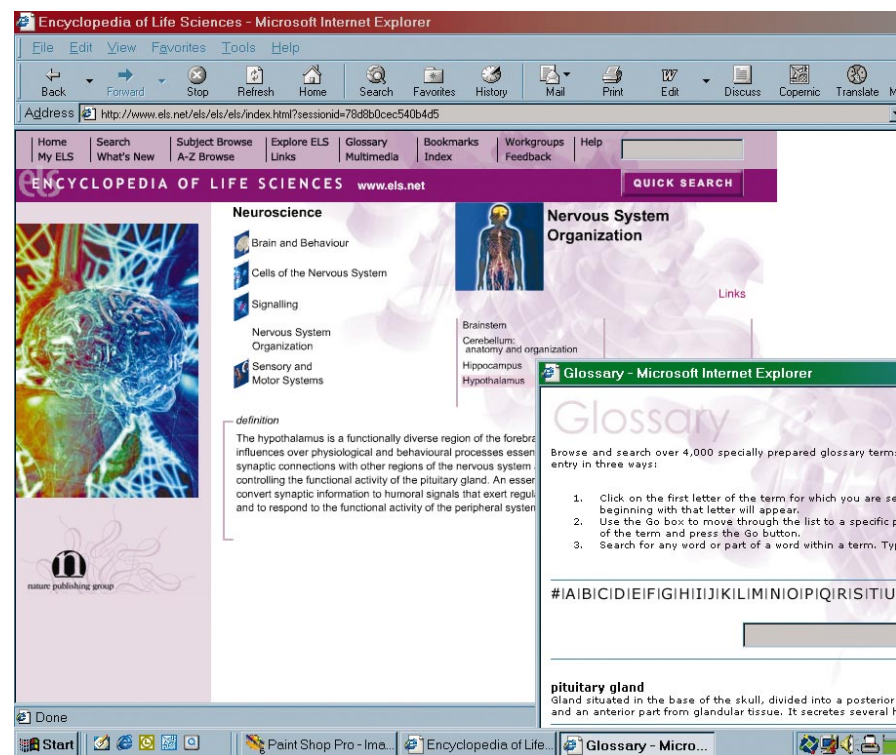
ca's executives that their reliable product could not be seriously challenged by CD-ROM encyclopaedias of possibly lower quality, set *Britannica* on a course that proved near-catastrophic. Not only were encyclopaedias on CD-ROM being given away with new computers, but also 'the computer' replaced 'the encyclopaedia set' as the major purchase made by parents anxious to assist in their children's education. A few years on, however, there is hope on the horizon: the new owners have embraced CD-ROM and DVD technologies, and have launched the post-modern phase of its existence — as a web portal for information seekers.

So, is the launch of a brand new encyclopaedia foolhardy or boldly innovative? With the backing of *Nature's* name for content excellence, and the fact that it has been designed primarily as an online resource, the *Encyclopedia of Life Sciences (ELS)* makes an interesting prospect.

There are several ways that we can get to the 3,000 articles by the commissioned authors — by keyword searching, or by browsing a subject or alphabetical list. Each *ELS* entry is graded as being either for non-experts or for the more life-science savvy; there is also a category of 'special essays' on particularly controversial or topical subjects. It would have been entertaining and useful to browse a list of these essays, but I couldn't find an obvious way to do this.

How to pitch the search? One aspect of a website as opposed to a book is that one cannot intuitively achieve a sense of the whole, although the promised *ELS* index should help users assess this more easily. After some experiment using the various article search fields ('title word' being the default), I found it best to search the full text, as would be the case for many search engines. A title search was much faster but often too restrictive, and the keywords seemed idiosyncratic — for example, an article entitled "Bioinformatics" is not found when "bioinformatics" is used as a key word.

Once the right questions are asked, the *ELS* world is at your feet. Of the tiny fraction of content familiar to me, I saw what I



Life online: ask the right questions, and the *ELS* world is at your feet.

expected, inspiring confidence. But the vista of the unknown was far more interesting.

By searching and navigating through copious "see also" hyperlinks to other *ELS* entries (some are not yet linked) and using the "related articles" function, I traversed from bread-and-butter articles on the central dogma and RNA structure to principles of geological time and Aristotle, stopping by methanogenesis and brain stems en route. The essays are a digestible size, well written and frequently well illustrated. For those that need more explanation of terms, there is a glossary, although it seems somewhat lonely in its own window, and could perhaps be more usefully deployed by linking it directly to appropriate content.

If the *ELS* is disappointing in any way, it is with respect to some of the technical aspects of the website. Although the web pages are beautiful, bandwidth is still an issue for most of us, and heavy use of graphics does not assist fast downloads. This, in combination with other features such as the use of frames, several pop-up windows and an occasionally temperamental search engine, often makes the *ELS* feel slow and clunky.

But perhaps most important, the *ELS* has

made only minimal use of hypertext linking to other online resources. The links from article citations to their abstracts are sparse, even for very recent articles from well-known journals that are certainly indexed by PubMed. Although stand-alone lists of links to useful websites are supplied, the few URLs cited in the articles are printed out and not hyperlinked.

Over and above these basic points are missed opportunities, for example, to link to database entries of genes discussed, author biographies, online teaching resources, or perhaps even to allow keyword-generated automatic searches of PubMed or the *Nature* journals.

Of course, with hyperlinks come maintenance issues, so perhaps these omissions were deliberate, or they are planned for the next phase of development. If so, then with some further engineering the *ELS* will graduate from being a useful stand-alone reference work to a true information portal, in the proud tradition of its ancestors. ■

Johanna McEntyre is at the National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 45 Center Drive, Bethesda, Maryland 20892, USA.



Poking around among the mammals

This melanistic variant of the checkered elephant shrew, *Rhynchocyon cirnei*, is just one of some 46,000 species catalogued and described in *The New Encyclopedia of Mammals* edited by David Macdonald (Oxford University

Press, £35). Lavishly illustrated, the encyclopaedia is a completely updated revision of its 1984 predecessor. The conservation status and future prospects of each species are also documented. ■

A new window on space

Revealing the Universe: The Making of the Chandra X-ray Observatory

by Wallace Tucker & Karen Tucker
Harvard University Press: 2001. 296 pp.
\$27.95

Ingo Lehmann

Revealing the Universe tells the exciting story of the birth of X-ray astronomy and the development of the Chandra X-ray Observatory, which was launched by the Space Shuttle Columbia in 1999. The observatory, previously known as the Advanced X-ray Astrophysics Facility, is one of NASA's four Great Observatories for space astrophysics, together with the Hubble Space Telescope, the Compton Gamma-Ray Observatory and the Space InfraRed Telescope Facility. Its unique imaging capability, high sensitivity and spectral resolution in the soft and hard X-ray wavebands enable us to investigate the nature of high-energy phenomena from such extraordinary objects as supernovae, neutron stars and black holes.

X-ray photons have such high energies that they interact strongly with the Earth's atmosphere, making it impossible to use ground-based instruments for observing astronomical X-ray sources. In 1962, Riccardo Giacconi, Herbert Gursky and colleagues, using a simple X-ray detector carried on a rocket, discovered the first cosmic X-ray source, Sco X-1 (in the constellation Scorpius), and the cosmic X-ray background, opening up X-ray astronomy as a new window on the Universe.

The success of the Uhuru X-ray satellite in the early 1970s led to calls for a large X-ray telescope that could provide data of almost the same quality as an optical telescope. Wallace and Karen Tucker describe the technological and political challenges from the first proposal for a large X-ray telescope in 1970, to the launch of the Chandra X-ray Observatory some 30 years later. They offer an inside view on the efforts by many people, including Harvey Tananbaum, Giacconi, Martin Weisskopf and Charlie Pellerin, to implement this enormous NASA project.

The book focuses more on the story of Chandra than on astrophysics, but it does present an overview of the first impressive results from the observatory's mission. One minor shortcoming is the lack of a mention of Europe's recent contributions to X-ray astronomy, in particular those made by the XMM-Newton observatory. ■

Ingo Lehmann is at the Astrophysikalisches Institut Potsdam, An der Sternwarte 16, 14482 Potsdam, Germany.

Science in culture

Spiritual shapes

Ernst Haeckel's 'art forms in nature'.

Martin Kemp

The "argument from design" held that the wondrous forms and mechanisms of the created Universe bear irrefutable witness to the existence of a supreme deity. Charles Darwin's theory of natural selection seemed to deal it a devastating blow. If, to use Ernst Haeckel's reasoning, there were no preconceived designs but only selection according to "natural choice", God could no longer be hailed as the direct designer of every individual cog in nature's clockwork. If God were to survive at all, 'He' needed to be located elsewhere in the temporal system through which things have come into being.

Haeckel was born in Potsdam in 1834 and

was the long-term inaugural professor of zoology at the University of Jena (1865–1909). He achieved a very particular and highly influential solution to the dilemma of being both an evolutionist and a deist. Famous for his formula that "ontogeny recapitulates phylogeny" and notorious for the eugenic implications of his ideas (see *Nature* 395, 447; 1998), he saw the apparently contradictory aspects of nature — organic and inorganic, spiritual and material, matter and energy — as the "attributes of the one underlying substance". This substance, at once material and psychic, was identical with God.

According to Haeckel, everything was infused with the substance's formative and psychic spirit, ranging from the "crystal soul" of inorganic materials, through the "cell soul" of the simplest

unicellular organisms, to the highest consciousness of mammals. In this context, God could be characterized as "the infinite sum of all natural forces, the sum of all atomic forces and all ether-vibrations". This spirit was not to be thought of as an abstract entity infused in the physical world from an exterior being, because "our scientific experience has never yet taught us of the existence of forces that can dispense with a material substratum". Haeckel reserved particular scorn for the "anthropomorphic representation of God", which "degrades this highest cosmic concept to that of a gaseous vertebrate".

The supreme popular manifestation of Haeckel's witness to natural design was his *Kunstformen der Natur* (art forms in nature), issued between 1899 and 1904 in ten instalments, each with ten plates of marvellous organisms, ranging from the unicellular radiolarians, on which he was particularly expert, to birds with patterned plumage and horned ungulates.

The coloured and monochrome plates parade a visual feast of natural symmetries. Haeckel consciously draws out the orderliness of natural design in his own highly skilled depictions, which have been visually pitched to come across with full beauty in Adolf Giltch's accomplished lithographs.

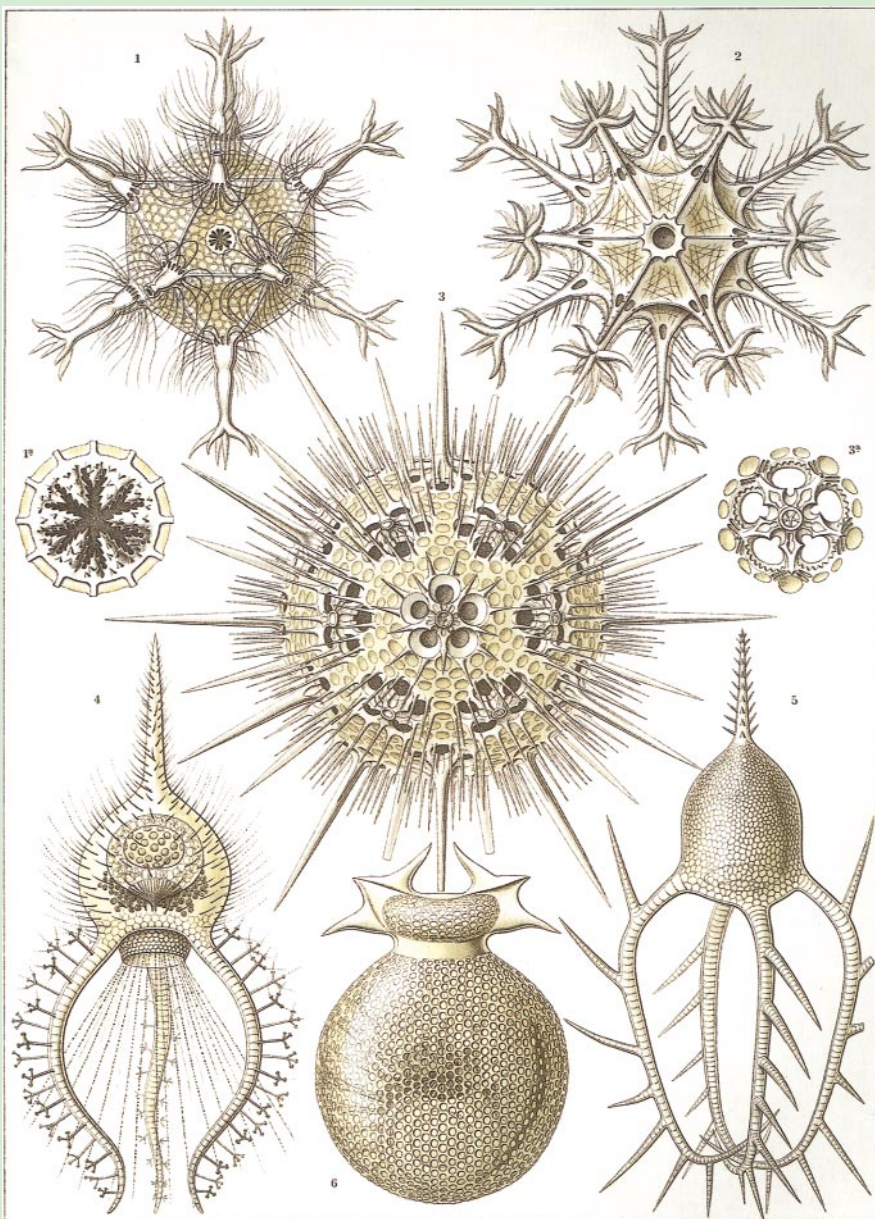
The superb plates of Haeckel and Giltch played an active part in the aesthetic of their time, adopting stylistic features of late-nineteenth-century design to enhance the formal wonders of the organisms, and in their turn influencing the mores of contemporary art. Aware that his images were beginning to affect the artists of his time, Haeckel claimed that "the discovery of countless beautiful forms of life ... [has] awakened a quite new aesthetic sense in our generation and thus given a new tone to painting and sculpture".

This "new tone" was nowhere more evident than in the international craze for art nouveau, or *Jugendstil*. In particular, the miracles of micro-design represented by the skeletons of the radiolaria, in which the "crystallography of organic forms" seemed most evidently disclosed, provided fertile inspiration for the extravagant organic symphonies of spikes, tendrils and plastic curves that became so fashionable around 1900. In so doing, Haeckel's plates infused a sense of visual joy into darwinian evolution that we have not entirely discarded today. ■

Martin Kemp is in the Department of the History of Art, University of Oxford, Oxford OX1 2BE, UK.

Haeckel's illustrations are on show in the *Turmalerie der Orangerie von Sanssouci*, Potsdam, until 15 October 2001.

Visualizations: *The Nature Book of Art and Science* is a collection of essays edited by Martin Kemp (published by Oxford University Press and the University of California Press; £20, \$35).



Haeckel's depiction of Phaeodaria, a class of unicellular radiolarians, on which he was especially expert.

Tales behind the tags

Physics is crowded with evocative phrases, but these alone cannot show the whole picture.

Philip Morrison

“What is it that human beings ultimately depend upon? We depend on our words. We are suspended in language. Our task is to communicate ... without losing the objective or unambiguous character [of what we say].” Thus spoke Niels Bohr, the most linguistic of oracles in twentieth-century physics. Yet the charm of memorable speech can and does mislead us all.

Take the tag ‘relativity’. It bears the memory of Einstein as a kind of catchword. But it was equally the work of that gifted First Physicist, Galileo Galilei, expressed in 1632 in a book of vivid conversation. His central character, Sagredo, recalls the first hours of a voyage out of Venice one calm and tranquil day. Once afloat, he wrote in his journal. Attentive Simplicio puts it well: “The gross motion from Venice ... was common to the paper, the pen, and everything else in the ship. But the small motions communicated to the pen by the fingers, but not to the paper, could leave a trace.” The two lines traversed by pen and paper as they were carried along on the ship were miles long, yet never so much as a foot apart. They thus differ subtly; it is the complex shifting distance between them that builds the record. Only relative motion counts locally, whenever the common motion is sufficiently uniform in its speed and direction. We all move as the Earth moves; we can view ‘gross motion’ in the sky or from it, but we do not feel it.

That was the first relativity. It was named ‘galilean relativity’, only after that famous paper was detonated in 1905 by a young Albert Einstein. By then, it was no longer planetary motion but electromagnetism that held puzzles. Only the relative motion between a magnet and a coil of wire matters to the dynamo; all agreed on this, including Einstein. Only the speed of light in space is not relative as other speeds are; it does not depend on whether the source of light or the detector moves, but only on their relative motion. By seeking the winds of the ether, physicists should have revealed the ‘gross motion’ of the Universe, but they certainly did not. No apt experiment, using any sensitive optical or electrical measurements, showed the ether drifting by.

Einstein explained this in six meticulous pages, using minor algebra. He boldly did away with the ether in his assumptions, and instead fixed for all observers the well-known absolute speed of light in empty space. All

The first relativity was named ‘galilean relativity’ only after that famous paper was detonated in 1905 by a young Einstein.

other speeds were relative, reckoned by a new calculation. This had extensive implications for measurements of time and motion — fundamental intervals, such as the duration of a watch tick, and space intervals, depend on the speed of the experimenter relative to the instruments he sets up. The closer a simple motion comes to light’s huge limiting speed, the more galilean relativity gives way. But relative motion is the basis of Galileo’s space in three dimensions, just as it is for Einstein’s space-time in four. Yet ‘relativity’ remains a tag that is associated with Einstein.

A more recent tag — which appears everywhere in vernacular cosmology, from comic strips to scientific papers — is the ‘Big Bang’. Our cosmology is rickety compared with the stringently tested 1905 ‘special relativity’ of Einstein. His ‘general relativity’, which describes curved space-time and incorporates gravitation, still frames our cosmology, the foundations of which are only two or three decades old. It ought to surprise us that the primal grand event, which is claimed to be the origin of all that there is, including space-time itself, has such a flippant name. The term was coined during a 1950 series on BBC radio by Fred Hoyle, a brilliant, fruitfully iconoclastic astrophysicist, whose firm intent then was to put down the naivety of the simple idea. (Sir Fred, aged 86, passed away in August this year.)

The Big Bang was based on a smooth extrapolation to the limit of cosmological equations (first written by Einstein in 1918 and significantly elaborated upon). In the simplest model, the Universe began at an implicit infinite limit, in which space was completely filled by very different matter from that which today makes up the Universe. Nowadays, the data from the sky favour a much more complex narrative featuring early staccato changes, called inflation, which have interrupted and indeed gave rise to the present long-standing expansion. The term Big Bang remains, confusingly, still in use for the wider observable



Seafaring story: a voyage from Venice was the setting for Galileo's description of relativity.

features that followed from a very much stranger and incompletely known past. We do not know what the beginning was like — or even if there was a beginning.

Bohr argued that physics concerns not what nature is, but rather what we can clearly say (and not say) about it. Science owes the outside world a clear, brief account of its views, but these accounts will never be clear until we go beyond mere buzz-words and start conveying real ideas, as Sagredo did. Perhaps we can enlist verbs to illuminate the nouns. Poetic, even ironic, tags catch on well, but name tags are just not enough. ■

Philip Morrison is in the Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

FURTHER READING

Galilei, G. *Dialogue Concerning the Two Chief World Systems* 2nd edn (trans. Drake, S.) 171–172 (Univ. California Press, Berkeley, 1967).
French, A. P. & Kennedy, P. J. (eds) *Niels Bohr: A Centenary Volume* 301–302, 305 (Harvard Univ. Press, Cambridge, Massachusetts, 1985).

The natural economy

Jeff Harvey

Towards the end of *On the Origin of Species*, Darwin pictures nature's riotous profusion of biological diversity occurring not as species in isolation, but with an interconnectedness he describes as an "entangled bank". Darwin was essentially describing an ecosystem, which science now defines as 'an integrated system involving the biological, chemical and physical components of a habitat or region'.

QUESTIONS THAT ECOLOGISTS MUST ADDRESS

- How do the cumulative biological and physical activities of individual organisms contribute to larger-scale ecosystem processes such as productivity, resilience and other measures of stability?
- How much can a given ecosystem be reduced in size, through loss of species, populations and guilds (see text), and still function effectively?
- For a given ecosystem, what are the most important functional components, and how much redundancy is there?
- How do direct and indirect interactions across ecosystem gradients (for example, above and below ground) determine the health and stability of the system?
- What are the short- and long-term ecological consequences of human alteration of the biosphere?

VICKY ASKEW

Healthy ecosystems are characterized by sustainable turnovers of energy, nutrients, organic matter and water, which remain stable over comparatively long periods of time. This homeostasis emerges from the collective functions of the system's biota, operating within an exceedingly complex, integrated network. Stability is maintained by functional redundancy, whereby multiple species have similar ecological roles, thereby allowing function to be maintained even in the face of change. In their aggregation over the entire biosphere, ecosystem processes are critically important in regulating global cycles of water, oxygen, nitrogen and carbon dioxide.

Humanity, like all of nature's bounty, is a product of evolution, and a dependence on and affinity for nature are undeniable components of the human psyche. Every human being lives within an ecosystem; towns, cities, farms and factories are freely leased to us by one or more ecosystems. We share ecosystems with literally billions of other organisms, most of which are wholly beneficial to us. Without this teeming array of biodiversity, life as we know it would not be possible. Yet many of us have become so insulated by technology within our artificial urban environments that we fail to appreciate the inputs and outputs of nature's economy upon which life itself—including our own—depends.

In the current age of artificial intelligence, human genomics, perceptual robotics, flow visualization, advanced astrophysics and other high-tech whizzery, it is both disturbing and ironic that our understanding of vital ecological processes is rudimentary. One of the greatest intellectual challenges facing the scientific community is to unravel ecological complexity and to understand better how ecosystems evolve, how they assemble themselves and how they function. The importance of this immense task is greater than it has ever been because of the drastic changes in the functioning of ecosystems caused by humans. Through continued, unrestrained economic development, we may push ecosystems beyond a critical threshold, beyond which they are unable to sustain themselves and, ultimately, us.

Over the past decade, ecologists from different specialized disciplines have accrued a wealth of data, which go some way to unravelling ecological complexity. But only partial headway has been made in understanding the relationship between biodiversity and ecosystem functioning. There has been little attempt to link the results of small-scale, mechanistic studies with research at the level of the ecosystem. The gap between macroscopic and microscopic processes must be closed, so that we

Ecosystems

A better understanding of existing data sets, combined with further research, is needed to predict present and future impacts of humanity on our global life-support systems.

can better understand how the complex biosphere has emerged from evolutionary pressures operating on small scales.

Moreover, most ecosystem experiments have been carried out using a small subsample of species that occupy one or two trophic levels in closed systems (microcosms), rather than on larger ranges of multitrophic interactions within an entire ecosystem or community. Admittedly, such blanket studies represent a gargantuan undertaking. But as long as species and guilds (groups of species that exploit resources in similar ways) are excluded from experimental programmes, the conclusions of these experiments will be open to doubt.

A better understanding of existing data sets, combined with further research, is needed to identify the present and future impacts of humanity on our global life-support systems, and to establish conservation priorities. But how can we achieve this, given the complexities? One way is to adopt a multidisciplinary approach, pooling and focusing resources and expertise on particular systems of which there is some historical knowledge. The efforts of ecologists who study ecosystems need to be combined with those who study individuals and populations.

Ecological science is crucial for policymakers attempting to mitigate the changes that humanity has inflicted on the biosphere. Without a better grasp of ecological processes, future generations will cause even more serious damage. The consequences for the planet are not worth contemplating. ■

Jeff Harvey is in the Department of Multitrophic Interactions, Centre for Terrestrial Ecology, PO Box 40, 6666 ZG Heteren, The Netherlands.

FURTHER READING

- Baskin, Y. *The Work of Nature* (Island, Washington DC, 1997).
- Jones, C. G. & Lawton, J. H. (eds) *Linking Species and Ecosystems* (Chapman & Hall, New York, 1995).
- Levin, S. A. *Fragile Dominion: Complexity and the Commons* (Perseus, Reading, Massachusetts, 1999).
- Maurer, B. A. *Untangling Ecological Complexity: The Macroscopic Approach* (Univ. Chicago Press, Chicago, 1999).
- Pimm, S. L. *The Balance of Nature? Ecological Issues in the Conservation of Species and Communities* (Univ. Chicago Press, Chicago, 1991).



Talk of genetics and vice versa

Steven Pinker

Does our ability to talk lie in our genes? The suspicion is bolstered by the discovery of a gene that might affect how the brain circuitry needed for speech and language develops.

"Man has an instinctive tendency to speak, as we see in the babble of our young children," wrote Charles Darwin¹ in 1871, "while no child has an instinctive tendency to bake, brew, or write." Darwin's observation has just been supported in a way he could not have dreamed of, with the discovery by Lai and colleagues² (page 519 of this issue) of a gene that is mutated in a disorder of speech and language.

The possibility that human language ability has genetic roots was raised about forty years ago by the linguist Noam Chomsky³ and the neurologist Eric Lenneberg⁴. Chomsky noted that language is universal, complex and rapidly acquired by children without explicit instruction. Lenneberg pointed out that a small number of children fail to display this talent and that such deficits sometimes run in families. Deficits of this kind are now called 'specific language impairment', an umbrella term for language disorders that cannot be attributed to retardation, autism, deafness or other general causes. Specific language impairment not only runs in families but is more concordant in identical than in fraternal twins, suggesting that it has a heritable component⁵. But the inheritance patterns are usually complex, and until recently little could be said about its genetic basis.

Then, in 1990, investigators described the 'KEs' — a large family, of several generations, in which half the members suffer from a speech and language disorder⁶. This disorder is distributed within the family in a manner that suggests it is caused by a dominant gene, or a set of linked genes, on an autosomal (non-sex) chromosome. The press referred to it as a 'grammar gene' (Fig. 1), while sceptics suggested that it merely lowers intelligence or makes speech unintelligible, or even that the disorder is nothing more than an artefact of a working-class dialect.

Extensive testing by psycholinguists, including Faraneh Vargha-Khadem, one of the authors of the paper in this issue², suggested that the disorder is more complex than either of these extremes^{7,8}. Affected family members do tend to score below average in intelligence tests (perhaps because verbal coding helps performance in a variety of tasks). But the language impairment cannot be a simple consequence of low intelligence, because some of the affected members score in the normal range, and some score more highly than their unaffected relatives.

And although the affected members have problems in articulating speech sounds (especially as children) and in controlled movements of the mouth and tongue (such as sticking out their tongue, or blowing on command), their language disorder cannot be reduced to a problem with motor control. They also have trouble identifying basic speech sounds, understanding sentences, judging grammaticality, and other language skills. For example, as adults they stumble at a task involving nonsense words that most four-year-olds pass with ease: completing sequences such as 'Every day I plam; yesterday I _____'.

In 1998 several of the authors of today's paper linked the disorder to a small segment of chromosome 7, which they labelled SPCH1 (ref. 10). Now, thanks to the discovery of an unrelated person known as CS, who has both a similar speech deficit to the KEs and a chromosomal translocation affecting the SPCH1 segment, Lai *et al.*¹ have narrowed the disorder down to a specific gene, *FOXP2*. In CS, this gene is disrupted by the translocation. In all the affected members of the KE family examined, but in none of the unaffected members, and in none of 364 chromosomes from unrelated, unaffected people, a single guanine nucleotide is replaced by an adenine. (The perfect contingency is in striking contrast to the now-you-see-it, now-you-don't correlations found in the first generation of searches for genes affected in behavioural disorders.) The authors propose that the nucleotide replacement results in substitution of the amino acid histidine for an arginine in one structure — the 'forkhead' domain — in the gene's protein product, presumably altering the protein's function.

Lai *et al.* present hints that *FOXP2* may have a causal role in the development of the normal brain circuitry that underlies language and speech, rather than merely disrupting that circuitry when mutated. *FOXP2* belongs to a family of genes that encode transcription factors (proteins that trigger the copying of genes into messenger RNAs), many of which have important roles in embryonic development. One of the defining features of proteins in this family is the forkhead domain, which contacts a target region in DNA, and it is this domain that is affected by the mutation in *FOXP2*. *FOXP2* appears to be strongly expressed in fetal



Figure 1 Genes and speech: a cartoonist's view of a 'language gene'. The identity of a gene affecting speech and language has been pinned down by Lai *et al.*², writing in this issue. (Reprinted with special permission, North America Syndicate.)

brain tissue (among other places), and its homologue is expressed in the developing cerebral cortex of mouse embryos. In both CS and the affected members of the KE family, only one copy of *FOXP2* is disrupted. So Lai *et al.* suggest that, at a critical point in fetal brain development, affected individuals have only half the normal amount of functioning transcription factor, which is not enough to control some aspect of early brain development.

Whatever the exact function of the gene turns out to be, the new work² has many implications. As a smoking gun for a genetic cause of one kind of language disorder, the discovery motivates the search for genetic causes of cognitive and learning disorders more generally, relieving the presumption of guilt from mothers (who are often still blamed for everything that goes wrong with their children). It also shows that just because a cognitive disorder has a genetic cause, it is not necessarily untreatable. The affected KE adults learned to compensate for their difficulty in generating complex linguistic forms by memorizing the forms

whole and by consciously applying rules they had been taught in language therapy¹¹. These and other strategies allow them to converse competently, although this has made life difficult for psycholinguists trying to work out the underlying disorder from the behaviour of affected adults.

If *FOXP2* really does prove necessary for the development of the human faculty of language and speech, one can imagine unprecedented lines of future research. Comparisons of the gene in humans to those in chimpanzees and other primates, and analyses of the types and patterns of sequence variation within the region of *FOXP2*, could add to our understanding of how human language evolved^{12,13}. An examination of the functions and expression patterns of the gene (and of other genes it might set off) in fetal and adult brain tissue could shed light on how parts of the human brain are prepared for their role in cognitive information processing.

The discovery of a gene implicated in speech and language is among the first fruits of the Human Genome Project for the cognitive sciences. Just as the 1990s are remem-

bered as the decade of the brain and the dawn of cognitive neuroscience, the first decade of the twenty-first century may well be thought of as the decade of the gene and the dawn of cognitive genetics. ■

Steven Pinker is in the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. e-mail: steve@psyche.mit.edu

1. Darwin, C. *Descent of Man* (John Murray, London, 1871).
2. Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. *Nature* **413**, 519–523 (2001).
3. Chomsky, N. *Language* **35**, 26–58 (1959).
4. Lenneberg, E. in *The Structure of Language: Readings in the Philosophy of Language* (eds Fodor, J. A. & Katz, J. J.) 579–603 (Prentice-Hall, Englewood Cliffs, NJ, 1964).
5. Stromswold, K. *Language* (in the press).
6. Hurst, J. A., Baraitser, M., Auger, E., Graham, F. & Norrell, S. *Dev. Med. Child Neurol.* **32**, 347–355 (1990).
7. Gopnik, M. & Crago, M. *Cognition* **39**, 1–50 (1991).
8. Vargha-Khadem, F. et al. *Proc. Natl Acad. Sci. USA* **92**, 930–933 (1995).
9. Pinker, S. *Words and Rules: The Ingredients of Language* (Basic Books, New York, 1999).
10. Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P. & Pembrey, M. E. *Nature Genet.* **18**, 168–170 (1998).
11. Ullman, M. T. & Gopnik, M. *Appl. Psychol.* **20**, 51–117 (1999).
12. Kreitman, M. *Annu. Rev. Genom. Hum. Genet.* **1**, 539–559 (2000).
13. Aquadro, C. in *Limits to Knowledge in Evolutionary Biology* (eds Clegg, M. T., Hecht, M. K. & MacIntyre, J.) 135–149 (Plenum, New York, 1999).

on the verge of science fiction such as the quantum computer.

Six years ago a new state of matter^{2–7} — the Bose–Einstein condensate (BEC), named after those who predicted its existence — was first created in a dilute gas of atoms. In a BEC, the usual energy distribution for an ensemble of particles no longer exists; all particles are forced to acquire the same energy. Furthermore, this energy is always the lowest allowed by quantum theory; it can be close but not equal to zero. A BEC contains up to ten million atoms, all at a temperature just above absolute zero (a few nanokelvin). In such a state, the macroscopic cloud of atoms has quantum features, which are distinctly different from those of the classical world we observe around us.

Until now, such clouds of ultracold atoms have only been handled from a distance. This is mainly because a BEC is so delicate that any contact with other atoms will destroy it. For this reason, BEC experiments are performed inside ultrahigh-vacuum chambers, providing an environment similar to that found in space. The clouds are trapped, manipulated and observed in magnetic, electric or light fields, which usually originate from sources outside the chamber, such as lasers or magnetic coils. The geometry of traps produced by these sources is therefore limited. A source close to the BEC could provide much tighter and more complex traps, but there were fears that the ultralow-temperature cloud would not survive in the presence of higher-temperature objects.

The achievement of Reichel and colleagues¹ in Munich — and the parallel work by C. Zimmermann's group in Tübingen⁸ — is to put the source of the trapping fields inside the ultrahigh-vacuum chamber, a few tens of micrometres away from the atom cloud. The experiments solve both of the

Bose–Einstein condensates

Mastering the language of atoms

Ron Folman and Jörg Schmiedmayer

Physicists can already make ultracold atoms perform quantum tricks in sophisticated magnetic and optical traps. But a fast route to trapping atoms on a microchip opens up new possibilities.

Atoms are the building blocks of all matter. They have a positively charged nucleus and their outer boundaries are defined by electron clouds. They remain electrically neutral, but the number of electrons governs their chemical properties. Atoms have long been studied and exploited

by mankind. Yet we are just now learning a whole new way of communicating with them. On page 498 of this issue, Reichel and colleagues¹ describe another step on this journey. Their achievement may result in new insights into the foundations of quantum theory, and lead to applications

Box 1 The atom-chip toolbox

Many of today's electronic devices are unthinkable without miniaturization. By similarly shrinking elements used in atom optics, such as atom traps, guides, mirrors, beam-splitters and interferometers, and by fabricating them using modern solid-state techniques (lithography) stemming from electronics and optics, physicists hope to achieve a similar level of control over atoms as they have over electrons and photons. The preparation,

manipulation and measurement sensitivity must reach a level at which quantum effects are dominant.

Why use atom chips? First, studying quantum behaviour requires the observed system to be isolated from its environment because any interaction would quickly destroy the delicate quantum effects. The neutral atom is an excellent choice in this matter — because it has no charge, it interacts

with its environment in a relatively weak way.

Second, chips offer a platform that is robust, scalable (it allows for arrays of traps, for example) and accurate. Together, atoms and chips make a powerful combination. Lithographic techniques can now create structures with length scales below 100 nm, which is smaller than the quantum-mechanical (de Broglie) wavelength of the cooled atoms, ensuring control at the quantum

level. The small size of the traps allows atoms to be positioned in individual sites separated by small distances, enabling them to interact in a controlled way. Because the atoms themselves are well localized (within 10 nm) they can be manipulated and detected by miniaturized light elements, such as micro-cavities and solid-state wave guides, which today can be fabricated on the same chip.

A long-term goal is to

fabricate everything on the same chip — from the light sources (micro-lasers) to the readout electronics — producing a truly integrated self-sufficient device. The hope is that such devices will do for quantum atom optics what integrated circuits did for electronics. Atom chips are already an outstanding research tool. Perhaps the day is not far off when they will also be household items, in clocks, communications and even computing. **R. F. & J. S.**

above problems: the proximity of the source enables more complex, tight and accurate manipulation of the BEC, and the BEC survives being close to the much warmer object. Moreover, the Munich researchers have taken it one step further by creating an atom 'conveyor belt' to move the BEC cloud around at will.

What's more, the Munich group showed that the initial formation of the BEC is also much simpler than in previous experiments, as less stringent ultrahigh-vacuum requirements are needed. This is because the tight traps they have created reduce the cooling time of the atoms by more than a factor of ten — a BEC can now be formed in less than a second. These short cooling times mean that any unwelcome atoms in the chamber have much less time to destroy the cloud.

How did they do it? Both experiments use a device called an 'atom chip'^{9–11}, in analogy to a computer chip (Box 1). In a conventional microchip, electrons move inside miniature 'traps' formed by the wires embedded in a solid-state object¹². In the atom chip, currents and charges also flow within wires patterned on the surface of a robust solid-state device. But unlike trapped electrons in a regular chip, here atoms hover a few micrometres above the surface of the chip — inside the invisible walls created by field potentials generated by the charges and currents inside the chip. Such surface traps help the atoms maintain their ultrahigh-vacuum environment. Researchers have already used atom chips to manipulate thermal atoms, but until now no one was sure it could be done with a BEC.

Now that we are able to trap and cool the atoms, we can control their position and velocity. We also know how to control their internal properties, such as the state of the electron clouds in the atom, known as hyperfine states. So all the relevant parameters can now be controlled and measured. This should encourage new experiments in atom manipulation (matter-wave quantum optics¹³), including chaos (using complex field potentials), nonlinearity (due to atom–atom interactions), entanglement (atom–atom correlations), atom–light interactions, low-dimensional physics and more.

Another issue in need of more insight is quantum decoherence. This is the process responsible for the classical features of our everyday world, despite its underlying quantum nature — it is what happens when a quantum state is destroyed. This elusive border between classical and quantum states has been a source of debate and confusion since the early days of quantum theory¹⁴. In atom-chip experiments, decoherence can be examined in complicated potentials and with carefully tailored environments. Furthermore, the question of surface-induced decoherence can be addressed in detail;

this is important because such decoherence may undermine the whole concept of the atom chip¹⁵.

What next? The atom chip might lead to miniaturized versions of highly accurate atomic clocks and acceleration sensors¹⁶, which are already used for precision measurements. Such tiny systems could be useful in navigation systems. The atom chip could also be integrated into quantum communication and encryption systems, which ensure information security. In quantum information, a 'qubit' is the quantum equivalent of a classical 'bit' of information. So, for example, the atom chip may enable the conversion of 'flying qubits' (photons that can travel along optical fibres) into 'storage qubits' — atoms that can be kept in a single location for a long time without changing their quantum state. A final example, and the most far-reaching, is the quantum computer, for which quantum theory predicts a new type of computing logic, able in some cases to outrun the present classical computers by many orders of magnitude in processing time^{17,18}. Once further advances are made on issues such as single-atom trapping and controlled entanglement, which are needed to do computations¹⁹, the atom chip could

turn out to be the obvious choice for building a quantum computer.

Ron Folman and Jörg Schmiedmayer are at the *Physikalisches Institute, University of Heidelberg, Philosophen weg 12, Heidelberg 69120, Germany.* e-mails: folman@physi.uni-heidelberg.de
schmiedmayer@physi.uni-heidelberg.de

1. Hänsel, W., Hommelhoff, P., Hensch, T. W. & Reichel, J. *Nature* **413**, 498–501 (2001).
2. Anderson, M. H. *et al. Science* **269**, 198–201 (1995).
3. Metcalf, H. & van der Straten, P. *Phys. Rep.* **244**, 203–285 (1994).
4. Ketterle, W. & van Druten, N. J. *Adv. Atom. Mol. Opt. Phys.* **37**, 181–236 (1996).
5. Chu, S. *Rev. Mod. Phys.* **70**, 685–706 (1998).
6. Cohen-Tannoudji, C. *Rev. Mod. Phys.* **70**, 707–719 (1998).
7. Phillips, W. D. *Rev. Mod. Phys.* **70**, 721–741 (1998).
8. Ott, H. *et al.* <http://arXiv.org/abs/cond-mat/0109322>
9. Folman, R. *et al. Phys. Rev. Lett.* **84**, 4749–4752 (2000).
10. Folman, R. *et al. Adv. Atom. Mol. Opt. Phys.* (in the press).
11. European ACQUIRE collaboration: <http://www.physi.uni-heidelberg.de/groups/acquire/summary.htm>
12. Buks, E. *et al. Nature* **391**, 871–874 (1998).
13. Berman, P. R. (ed.) *Atom Interferometry* (Academic, San Diego, 1997).
14. Giulini, D. *et al. Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer, Berlin, 1996).
15. Henkel, C. & Wilkens, M. *Europhys. Lett.* **47**, 414–420 (1999).
16. Peters, A., Chung, K. Y. & Chu, S. *Nature* **400**, 849–852 (1999).
17. Bouwmeester, D., Ekert, A. & Zeilinger, A. (eds) *The Physics of Quantum Information* (Springer, Berlin, 2000).
18. Bennett, C. H. & DiVincenzo, D. P. *Nature* **404**, 247–255 (2000).
19. Zoller, P. <http://th-physics.uibk.ac.at/qo/zoller/PZListOfPublications.html>

Bacterial genomics

A plague o' both your hosts

Stewart T. Cole and Carmen Buchrieser

The genome of the bacterium that causes plague is highly dynamic, and scarred by genes acquired from other organisms. Does this explain its ability to kill both mammals and insects?

The course of human evolution has been shaped by three major factors: natural disasters, wars and infectious diseases such as tuberculosis, syphilis, typhus and cholera. In the past few years the darkest secrets of the bacteria responsible for each of these diseases have been unveiled by genomics, and on page 523 of this issue¹ Parkhill and colleagues describe the latest subject of this approach — *Yersinia pestis*, the bacterium that causes plague. Ten years ago, at the height of Operation Desert Storm, the threat of biological weapons was ever present (and today that threat has reared its head again, in the form of bioterrorism). Foremost among these weapons was *Y. pestis*, a formidable airborne adversary that rapidly kills humans if left to its own devices.

There have been three plague pandemics, which were collectively responsible for the loss of 200 million lives. The second of these left an indelible mark on world history: over one-third of the population of Europe succumbed to the disease between 1347 and 1350. There was, however, some hidden benefit for the survivors and their descen-

dants, as the pandemic contributed to the elimination of leprosy from the continent by removing the reservoir of the causative organism, *Mycobacterium leprae*². At the time, leprosy patients were sequestered in leprosaria run by various religious orders, and neither the carers nor the cared-for were spared by the Black Death. Intriguingly, as we will see below, there are many parallels between the genomes of the plague and leprosy bacteria. During the third pandemic, which spread from south China in 1894, Alexandre Yersin isolated the bacterium that now bears his name, and a little later Paul-Louis Simond showed that fleas are involved in transmitting plague.

Yersinia pestis has an interesting yet complex natural history involving a mammalian reservoir and an insect vector (Fig. 1, overleaf). It is primarily a rodent pathogen, with a predilection for rats, and is usually transmitted to humans by fleas following the death of the rodent. The bacterium then spreads from the initial site of infection, the flea-bite, to the lymph nodes, where local replication causes a swelling, or bubo, lead-

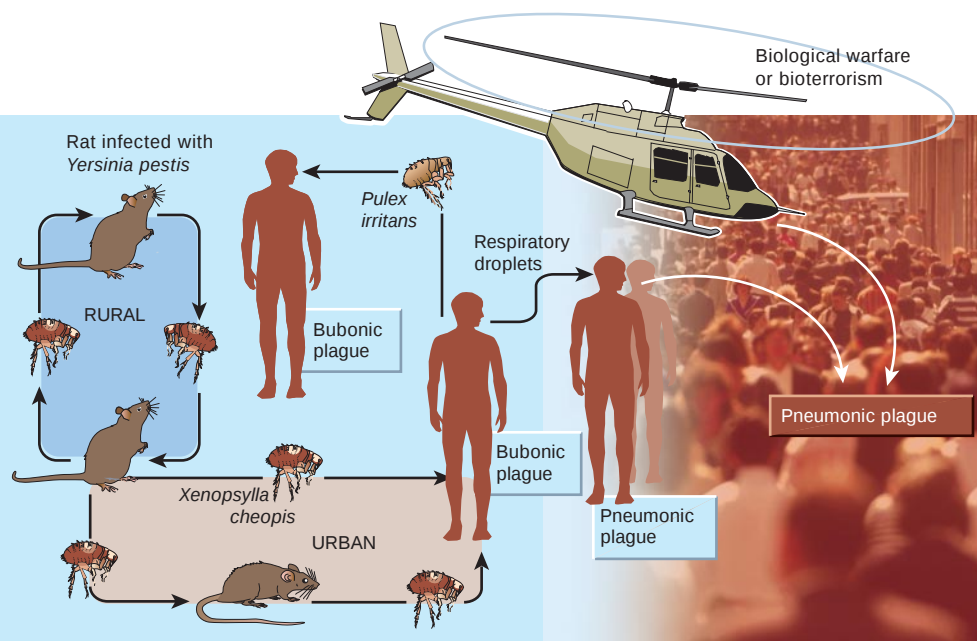


Figure 1 How the plague is transmitted. The causative organism, the bacterium *Yersinia pestis*, has a rodent reservoir. The rodents' fleas, such as the oriental rat flea, *Xenopsylla cheopis*, acquire *Y. pestis* from a meal of infected blood, and transmit the bacterium primarily to other rodents or to humans, causing bubonic plague in people. Human-to-human transmission can also take place, through the human flea *Pulex irritans*. Pneumonic plague is less frequent but even more severe; it is transmitted from person to person through respiratory droplets, or even by artificially generated aerosols, containing *Y. pestis*.

ing to bubonic plague. If the lymphatic system becomes overwhelmed, widespread dissemination and rapidly fatal blood poisoning result, with the bacterium spreading to all the main organs including the lungs. The disease then spreads in the air in droplets, culminating in highly contagious pneumonic plague, which can bring death in under three days. These properties make *Y. pestis* one of the most feared agents of biological warfare or bioterrorism.

Yersinia pestis is what's known as a bacillus — it is rod-shaped — and it belongs to a group of bacteria called the Enterobacteriaceae. Most of these bacteria are associated with animals and humans, and many are harmless. So what made *Y. pestis* such an aggressive pathogen and how, uniquely for this group, did it adapt to life in a blood-sucking insect? Parkhill *et al.*'s analysis¹ of the complete genome sequence of C092 — a *Y. pestis* strain isolated recently from a person who died of pneumonic plague — provides some fascinating answers.

The genome comprises a chromosome of 4.65 million base pairs, as well as three plasmids (circular DNA molecules) of some 96, 70 and 9.6 kilobases. As Parkhill *et al.* discover, this genome displays unusual fluidity: there is compelling evidence for the inversion of chromosomal segments and for gene acquisition and decay. Repeated sequences are abundant (making up about 3.7% of the total) and, as in the leprosy bacillus, these serve as sites for 'homologous recombination' events, which remodel the genome³.

By investigating the bias towards guano-

sine and cytosine nucleotides (the G/C skew)⁴ in different parts of the genome — a genomic parameter that highlights compositional irregularities — the authors detected three unusually large segments that had inverted or translocated. Analysis of the inversion endpoints indicated that both possible orientations could be found, in unequal proportions, in C092 samples, as well as in other strains, implying that several different chromosomal configurations can exist within the same population. As the expression of bacterial genes is influenced by their orientation with respect to the direction of DNA replication⁵, this raises the possibility that differences in strain virulence could stem from differences in genome arrangement.

Using the G/C skew and other bioinformatics tools, Parkhill *et al.* also identified 21 regions of the chromosome that show characteristics of pathogenicity or adaptation islands, probably acquired from other organisms through horizontal transfer. These islands reflect the twin hosts of *Y. pestis*. They code not only for molecules implicated in infecting mammalian cells — such as adhesin proteins, two new siderophores (iron chelators), and secretion systems — but also for functions that might enable *Y. pestis* to colonize, and transit through, fleas. For example, several genes encode potential insecticides, and a baculoviral-like enhancin protein, that could damage the insect midgut. As in gastrointestinal pathogens, genes that code for factors involved in virulence are either chromoso-



100 YEARS AGO

There is only one class of zoologist that I would wish to blot out, and that is the class whose reckless naming of new "species" and "varieties" serves only to extend the work and the tables of the conscientious synonymy hunter. Other than this all classes will contribute to the advancement of the science. No doubt there are unlabeled [*sic*] species and no doubt they must, as things are, be named. And no doubt genera and families must be "revised" and some groups split up and others lumped. So welcome to the old-fashioned systematist, though his day be short, and may he treat established genera gently. No doubt there are types of animals of whose structure we are woefully ignorant; no doubt we need to know their internal anatomy in great details. So welcome to the zootomist in this new century, and may he invent fewer long names for new organs. No doubt there are groups of whose relationships we know little, and which have been buffeted about from one class to another in a bewildering way. We need to have their places fixed. So welcome to the comparative anatomist and the embryologist, and may their judgment as to the relative value of the criteria of homology grow clearer. No doubt our knowledge of inheritance and development will be immensely advanced by the further study of centrosomes, asters and chromosomes. Welcome, therefore, to the cytologist, and may he learn to distinguish coagulation products and plasmolytic changes from natural structures. All these subjects have victories in store for them in the new century. To neglect them is to neglect the foundations of zoology.

From *Nature* 3 October 1901.

50 YEARS AGO

No scientific survey, however, would be complete without a visit to Sprotborough, near Doncaster, for here was Sprotborough Hall, the home of Sir Godfrey Copley, Bart. (d. 1709), who by his will left £100 to the Royal Society for "improving natural knowledge"... His £100 was allowed to lie idle for some time, but awards were made in 1731, 1732 and 1736, and then the money was used to found the Copley Gold Medal, now the oldest and most famous of prizes in the world of science... In 1749 the Medal was given to John Harrison, a native of Wragby, near Wakefield, for his "Curious Instruments invented and made by him for the exact Mensuration of Time".

From *Nature* 6 October 1951.

mal or found on plasmids that may have been acquired.

This microorganism is closely related to the food- and water-borne gastrointestinal pathogen *Yersinia pseudotuberculosis*, whose partial genome sequence was of help to Parkhill *et al.* in interpreting the *Y. pestis* genome. Although it is thought that *Y. pestis* might have evolved from *Y. pseudotuberculosis*⁶, they have profoundly different pathogenic strategies. This is reflected in the fact that many *Y. pseudotuberculosis* genes required for adhesion, mobility and colonization of the gut have decayed in the plague bacillus. Further genome comparisons will help to unravel other functional differences.

Gene decay in *Y. pestis* is also seen in the fact that it contains 149 pseudogenes, representing some 4% of the genome. This is a relatively modest proportion compared with the leprosy bacillus, in which over 50% of the genome has decayed, yet it is a firm indication that reductive evolution has commenced⁷. This process reflects adaptation to a new host and perhaps even the gradual

transition to a lifestyle in which the bacterium abides, in a symbiotic partnership, inside its host's cells. It might also herald the slow disappearance of the plague bacillus. Evolutionary time is long, however, and meanwhile the information provided by the *Y. pestis* genome sequence⁴ should be applied to ensure that plague does not re-emerge, and that one of the potential weapons of bioterrorism can be neutralized. ■

Stewart T. Cole is in the Unité de Génétique Moléculaire Bactérienne, and Carmen Buchrieser is in the Laboratoire de Génomique des Microorganismes Pathogènes, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France.

e-mails: stcole@pasteur.fr

cbuch@pasteur.fr

1. Parkhill, J. *et al.* *Nature* **413**, 523–527 (2001).
2. Girard, G. *Bull. Soc. Pathol. Exot. Filiales* **68**, 33–37 (1975).
3. Cole, S. T. *et al.* *Nature* **409**, 1007–1011 (2001).
4. Lobry, J. R. *Mol. Biol. Evol.* **13**, 660–665 (1996).
5. Brewer, B. J. *Cell* **53**, 679–686 (1988).
6. Achtman, M. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 14043–14048 (1999).
7. Buchrieser, C. *et al.* *Infect. Immun.* **67**, 4851–4861 (1999).

Global change

Chill taken out of the tropics

Lee R. Kump

A paradox in palaeoclimatology has been the apparent existence of a cool sea surface in the tropics under conditions of high CO₂ in the atmosphere. It looks as if that paradox has been resolved.

Over the next few centuries fossil-fuel burning is likely to drive atmospheric CO₂ to perhaps six times the pre-industrial level. To see what might happen in such circumstances, geologists have turned to the ancient geological record in search of similarly extreme 'greenhouse' conditions. Studies of warm intervals in Earth history have linked the warmth to high concentrations of CO₂, but have also produced an unexpected result: although temperatures at high latitudes increased markedly at such times, tropical temperatures seem to have been similar to or less than today's. This puzzling observation is in direct conflict with state-of-the-art climate modelling of the response of the climate system to increased atmospheric CO₂ — although high-latitude regions warm more in such models, they predict considerable warming at all latitudes.

On page 481 of this issue, Pearson and colleagues¹ show that this 'cool tropics paradox' is more apparent than real, reflecting the difficulty of avoiding the often-subtle biases of diagenesis — the alteration of sediments after they were deposited — in the geological record. Pearson *et al.* argue that, in the past, sea surface temperatures in the tropics may have been much warmer, not cooler, than today's, and may actually have

been detrimental to temperature-sensitive organisms such as corals.

The controversy centres on palaeoclimatic reconstructions derived from oxygen isotopic compositions of the calcareous shells of foraminifera, which provide proxy information on ocean temperatures. Foraminifera are protozoa that are abundant in both modern and ancient oceans. Shells of both planktonic (free-floating) and benthic (bottom-dwelling) foraminifera accumulate on the sea floor, along with the organic, calcareous and siliceous remains of other organisms and mineralogical remains from continental weathering.

In principle, the oxygen isotopic composition of the mineral calcite that makes up the foraminiferal shell (reported as $\delta^{18}\text{O}$, the ratio of $^{18}\text{O}/^{16}\text{O}$ relative to a standard, in parts per thousand) reflects the temperature and isotopic composition of the waters in which the shells formed. The $\delta^{18}\text{O}$ of the ocean is fairly uniform geographically, and is thought to vary through time principally as the result of the growth and retreat of continental ice sheets. So, during warm intervals of Earth history, when continental ice sheets were small, variations in foraminiferal $\delta^{18}\text{O}$ primarily reflected spatial and temporal variations in temperature. Tropical planktonic

foraminifera construct their shells at or near isotopic equilibrium with the warm waters in which they live. But when they die the shells settle to the sea floor, where water can be 10–30 °C cooler and corrode the shells. Subsequent recrystallization on or below the sea floor tends to shift the isotopic composition to reflect the cooler temperatures there.

Investigators have worried about diagenetic artefacts for decades, and strict criteria have been adopted to weed them out. Careful application of these criteria has lessened the severity of the cool-tropics paradox in the past few years, but there have still been considerable mismatches between climate-model predictions and the proxy temperature record from $\delta^{18}\text{O}$ of foraminiferal shells. Only last year, for instance, Crowley and Zachos² articulated the prevailing view that



Figure 1 Fossil thermometer — the shell of *Cribrohamkenina inflata*, a foraminiferan from around 34.6 million years ago, viewed through the scanning electron microscope. Pearson *et al.*¹ carried out isotope analyses of very well-preserved examples of species such as this in their re-assessment of past sea surface temperatures in the tropics under high-CO₂ conditions. They conclude that the tropics were much warmer than previously estimated, in line with the expectation that higher levels of CO₂ in the atmosphere result in higher temperatures at all latitudes because of the greenhouse effect. (Micrograph courtesy of P. N. Pearson. The specimen is about 1 millimetre long.)

"there is little robust geological evidence indicating that tropical SSTs [sea surface temperatures] increased as CO₂ increased".

Pearson *et al.*¹ suspected that even the strict diagenetic criteria might be insufficient to prevent skewing of the temperature record obtained from foraminifera. So they sampled some rare but exceptionally well-preserved sediments from two warm intervals of Earth history: from the Eocene epoch (about 50 million years ago) and the Late Cretaceous period (about 70 million years ago). Only those foraminiferal shells that appeared pristine at the microscopic scale were subjected to isotopic analysis (Fig. 1). According to the authors, most previous analyses of planktonic foraminifera have been on samples that had undergone significant microscopic recrystallization (the microscopically porous nature of planktonic foraminiferal shells makes them particularly susceptible to diagenesis).

The authors then converted these oxygen isotopic compositions to temperatures, using established empirical relationships and a correction for the estimated amount of water stored in continental ice sheets at times in the past (in general it was much smaller than today). The resulting estimates of tropical temperatures (28–32 °C) are considerably higher than previous figures based on planktonic foraminifera (15–20 °C); and they are consistent with studies^{3,4} of mollusc shells, which are more resistant to diagenesis than those of foraminifera. Tropical sea surface temperatures may actually have been much higher than today's 25–27 °C, and were perhaps high enough to stress corals and other temperature-sensitive organisms. This conjecture is supported by the displacement of reef-building corals to higher latitudes during these warm intervals^{5,6}.

Much palaeoclimatological research is aimed at evaluating the link between high levels of atmospheric CO₂ and global warmth. The new data and analysis by Pearson *et al.*¹ seem to solve a puzzle that has troubled people in this field for years. But isotope geologists have invested considerable effort in previous analyses of planktonic foraminifera, and they will require a thorough evaluation of the new results. It needs to be confirmed that continental runoff of isotopically light fresh water has not affected the isotopic composition of the samples collected by Pearson *et al.*, a possibility that the authors acknowledge. And other samples previously thought to be well preserved will have to be re-checked to see if they were significantly affected by microscopic recrystallization.

But if Pearson and colleagues' conclusions¹ are supported by further work, they could revolutionize palaeoclimatology. They could shift the focus of sample recovery to the clay-rich continental-margin environments that favour preservation, while removing an impediment to climate modellers who have struggled with the cool-tropics paradox

for over a decade. With a new era of ocean drilling on the horizon, and the need for improved abilities and confidence in climate-change prediction never greater, this discovery couldn't have come at a better time. ■

Lee R. Kump is in the Department of Geosciences and the Astrobiology Research Center, 535 Deike Building, Pennsylvania State University, University Park, Pennsylvania 16827, USA.

e-mail: lkump@psu.edu

1. Pearson, P. N. *et al.* *Nature* **413**, 481–487 (2001).
2. Crowley, T. J. & Zachos, J. C. in *Warm Climates in Earth History* (eds Huber, B. T., Macleod, K. G. & Wing, S. L.) 50–76 (Cambridge Univ. Press, 2000).
3. Andreasson, F. P. & Schmitz, B. *Paleoceanography* **13**, 183–192 (1998).
4. Kobashi, T., Grossman, E. L., Yancey, T. E. & Dockery, D. T. *III Geology* (in the press).
5. Johnson, C. C. *et al.* *Geology* **24**, 376–380 (1996).
6. Kiessling, W. *Geology* **29**, 751–754 (2001).

Neurobiology

Developing order

Rick Livesey and Connie Cepko

Different types of nerve cells are made in a particular sequence during development. But how? Studies of fruitflies reveal a temporal order in the expression of the genes that regulate these decisions.

The construction of a nervous system during development depends on two main steps: making the right cells in the right places at the right time, and wiring those cells together to produce functional circuits. A paper by Isshiki and colleagues¹, published in *Cell*, takes us considerably further towards understanding the first of these processes. The authors have found that a set of key cellular proteins in the fruitfly *Drosophila melanogaster* are expressed in an order such that the right neural cell types are made at the correct time. This is a groundbreaking finding for developmental biology, with implications for understanding how the genome is dynamically deployed to generate a huge diversity of neuronal types.

Development of the nervous system does not depend on the generation of a large population of identical, bland cells that are then recruited to make different neural circuits. Instead, distinctly different types of neurons are made in different parts of the nervous system in a particular order during development. But all neurons in a given lineage are produced from the same pool of dividing cells — known as stem and progenitor cells — so how does this stereotyped order come about? At any time in development, two interacting forces are thought to combine in a progenitor or stem cell before it divides, conferring particular fates on the two progeny cells. These forces are extracellular signalling molecules, such as growth factors,

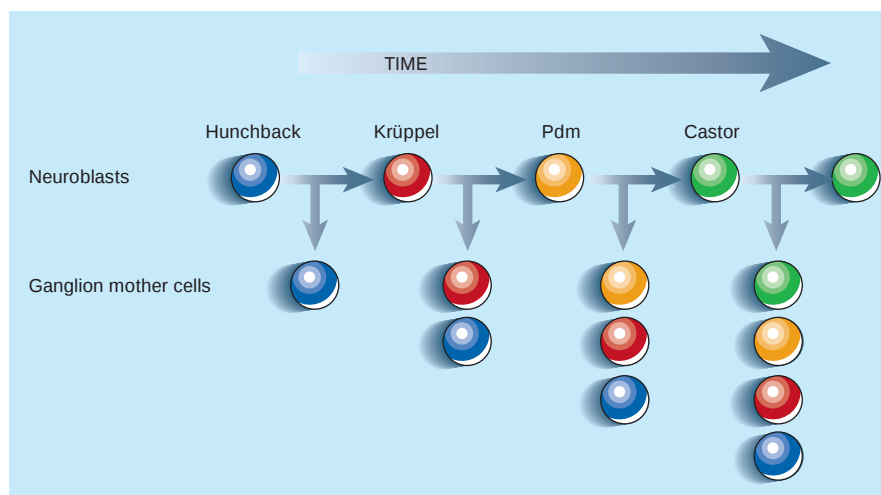


Figure 1 Controlling neuronal fate in fruitflies, according to Isshiki and colleagues¹. From left to right, dividing (mitotic) progenitor cells (neuroblasts) express four transcription factors — Hunchback, Krüppel, Pdm and Castor — in a temporal sequence that regulates the order in which neurons are produced. Each time a neuroblast divides, it gives rise to a new neuroblast and a ganglion mother cell, which itself divides to generate two postmitotic neurons (not shown) that maintain expression of the factor expressed by the progenitor before it divided. Neuroblasts change the transcription factors that they express in a cell-cycle-dependent manner, and over developmental time each neuroblast generates a series of ganglion mother cells, each expressing a different transcription factor. Although not shown here, some neuroblasts express two of the factors in the sequence at the same time, before going on to express only one.

and factors intrinsic to the progenitor or stem cell. A key role of the intrinsic factors appears to be to dictate a property referred to as 'competence' — that is, which subset of neuronal types can be generated from each progenitor or stem cell in response to extracellular signals.

As the competence of progenitors changes over time in many organisms, it is thought that the intrinsic factors that control competence also change². But there has been a lack of data to describe both the intrinsic influences and the motor that drives change. Obvious candidates for the intrinsic factors are transcription factors, proteins that could switch on and off particular programmes of gene expression to control the identities of different neurons. But the hunt for such factors was fruitless — until now. The paper by Isshiki and colleagues¹ has provided a welcome breakthrough in this field.

The authors¹ have extended previous studies^{3,4} to find a series of four unrelated transcription factors that are expressed transiently and sequentially in dividing neural progenitors (referred to as neuroblasts in *Drosophila*), but permanently in the progeny generated at the time that each factor is expressed (Fig. 1). So, for example, the first two factors in the sequence — Hunchback and Krüppel — were found to be both necessary and sufficient to generate neural cell types that maintain the expression of those factors. If the authors removed the function of either Hunchback or Krüppel from neuroblasts, then the neurons that normally express these factors were not produced. The generation of later neurons was completely undisturbed however, indicating that the steps controlled by each of these factors could simply be skipped without affecting later development. Conversely, overexpression of Hunchback or Krüppel forced neuroblasts to make more of the cell types that express these factors, but at the expense of cell types born later; fewer of the later cells in the sequence were made.

Although these are fascinating findings in their own right, Isshiki *et al.*'s most striking observation was that these factors are expressed in the same temporal order in completely different neural lineages at different times in development. Furthermore, the same sequence of factors is used to generate different cell types in each lineage. For example, in one lineage the first cell type produced is a motor neuron, while in another the first cell is a glial cell, but in both lineages the production of these first-born cells depends on Hunchback. The authors may have uncovered elements of a fundamental mechanism used in many parts of the nervous system, and perhaps in other tissues, to regulate cell-fate choices in a temporal sequence.

As if this were not enough, Isshiki *et al.* have further findings that provide insight

into the mechanisms regulating these intrinsic changes. They show that halting the cell-division cycle prevents progenitors from changing their expression of transcription factors. Moreover, releasing the cell-cycle brake at a later time results in progenitors that express the next factor in the series, rather than playing catch-up by skipping a factor and expressing the more temporally appropriate one. This suggests that the changing behaviour of progenitors is linked to cell-cycle progression rather than absolute time.

Predictably, there are several vertebrate homologues of each of these factors. It will be interesting to see if any of them are expressed in a manner that suggests they have a similar role in the vertebrate nervous system — particularly in the retina and cerebral cortex, in which progenitors are known to pass through different competence stages^{5,6}. Such studies will also reveal whether the changes in competence of vertebrate progenitors, and the temporal changes in *Drosophila* neuroblasts, are in fact similar cellular processes.

These findings open up several other avenues of further research. For example, how do these widely expressed transcription factors — which are also involved in another developmental process, mesoderm segmentation — confer specific cell fates on neurons? Do these factors, acting with others, control particular transcriptional networks in progenitors to regulate cell fate, and in newly generated neurons to regulate neu-

ronal identity? And what are the components of these networks in different cell types? The availability of whole-genome microarrays for *Drosophila*, which allow rapid analysis of gene-expression patterns in different circumstances, makes these questions tractable.

Another immediate problem is how cells integrate the activity of these factors with the information from extrinsic cues, brought into the cell by the signal-transduction machinery. And finally, it remains to be seen how neuroblasts use the cell cycle to control the sequential expression of transcription factors, and how these factors in turn interact with both the cell-cycle machinery and the well-described 'neurogenic' signalling pathways to regulate cell fate in *Drosophila*. ■

Rick Livesey is at the Wellcome/CRC Institute of Cancer and Developmental Biology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK.

e-mail: rick@welc.cam.ac.uk

Connie Cepko is in the Department of Genetics and Howard Hughes Medical Institute, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02113, USA.

e-mail: cepko@genetics.med.harvard.edu

1. Isshiki, T., Pearson, B., Holbrook, S. & Doe, C. Q. *Cell* **106**, 511–521 (2001).
2. Cepko, C. L., Austin, C. P., Yang, X., Alexiades, M. & Ezzeddine, D. *Proc. Natl Acad. Sci. USA* **93**, 589–595 (1996).
3. Kambadur, R. *et al.* *Genes Dev.* **12**, 246–260 (1998).
4. Brody, T. & Odenwald, W. F. *Dev. Biol.* **226**, 34–44 (2000).
5. Desai, A. R. & McConnell, S. K. *Development* **127**, 2863–2872 (2000).
6. Belliveau, M. J. & Cepko, C. L. *Development* **126**, 555–566 (1999).

Gravitational physics

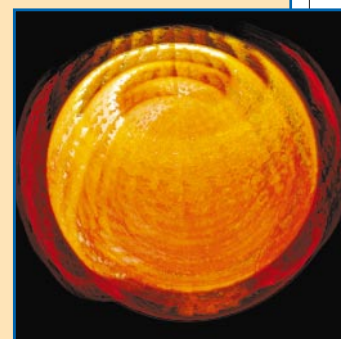
Black hole blockbuster

What happens when one black hole hits another black hole? Does it get more black? Seriously, though, such questions are being tackled by scientists seeking to detect gravity waves. These invisible waves are ripples in the fabric of space-time, predicted by Einstein's theory of relativity, but that remain undetected. They are thought to be emitted in copious amounts by colliding black holes, and a new simulation by researchers in Germany (J. Baker *et al.* *Phys. Rev. Lett.* **87**, 121103; 2001) gives an advance preview.

Collisions between other astronomical giants, such as galaxies, produce light and other radiation, but black-hole collisions generate only gravity waves. Black hole binaries are

thought to emit gravity waves all the time, but only when they collide are the waves strong enough to be detected on Earth. Three detectors are expected to start collecting data soon: the US LIGO and German–British GEO600 projects in 2002, and the Italian–French VIRGO detector in 2003.

To fully simulate the merger of two black holes, the German team merged — appropriately enough — two different approaches for calculating what might happen before and after the collision. In the computer-generated image shown here, spherical shells of intense gravity waves move outwards from the centre of the collision. The authors estimate that 3% of the total mass of the black holes is released as energy by



the collision — higher than expected.

These calculations should provide experimenters with a rough estimate of what to look out for, and guide more advanced simulations that take into account, for example, the likelihood that the black holes are also spinning. Prepare for a glimpse of the darkest corners of the Universe. Sarah Tomlin

W. BENDER, AEI/ZIB

Suspended by sound

E. H. Brandt

Ultrasound waves can levitate heavy balls of tungsten. This contact-free method of keeping items suspended in the air can be applied to the investigation and processing of new materials.

It is well known that radiation can exert a force. The solar wind, for example, is caused by sunlight blowing away microscopic dust particles, and its force on a sunbather is equal to the weight of a fly. It is probably less well known that sound waves also exert a force. The radiation force on a tiny sphere from a travelling sound wave is weak, proportional to the sixth power of the ratio of the sphere radius to the wavelength of the sound. But it can become more substantial — proportional to the third power of this ratio — in a standing sound wave, of the sort that occurs near reflecting walls. It has previously been shown that a siren operating at 3,260 Hz and an appropriate reflector can freely levitate drops of liquid and bubbles¹, or even a steel ball 1 cm in diameter². As Xie and Wei³ show in *Applied Physics Letters*, the levitating force of ultrasound (which has a frequency too high for humans to hear) can be enhanced even further by carefully designing the shape of the reflector. This allows them to levitate balls of high-density tungsten.

The strongest acoustic forces are exerted by the standing sound waves formed inside an almost closed box. But the 'single-axis geometry', in which a concave circular reflector faces a vibrating cylinder that emits ultrasound at a frequency of 16.7 kHz (Fig. 1), is more convenient and provides easy access to the levitating samples. Such levitators can be used to simulate the microgravity conditions of space in a terrestrial laboratory. Conversely, they are used to position samples

in space experiments, where there is no gravity. They have also been used to process materials where it is important to avoid the contamination of samples by the container wall — for example, for growing ice particles⁴, undercooling liquids far below their freezing point⁵, and heating liquid crystals⁶ and ceramics⁵ to high temperatures.

Xie and Wei³ have managed to increase the force and stability of single-axis acoustic levitation, and so were able to levitate balls of tungsten, which has a density of 18.9 g cm⁻³. They achieved this by optimizing the diameter of the vibrating cylinder and the distance, radius and curvature of the concave reflector. To understand the relationship between these geometric parameters and the observed enhancement, the authors developed a detailed model.

They began by computing the sound wave generated in their levitator from hydrodynamic theory to give the time-averaged mean square pressure and mean square velocity of the vibrating air at each point in space. They inserted these values into a general expression for the time-averaged potential of the acoustic radiation force acting on a small rigid sphere. This expression was first derived by Lev Gor'kov in an ingenious paper⁷ in 1962 (see also the discussion and application of this formula in ref. 8). It says that the sphere is attracted to regions with large air velocities and repelled by regions with high pressures. This seems intuitive — the sphere avoids places at high pressure and prefers places with large nega-

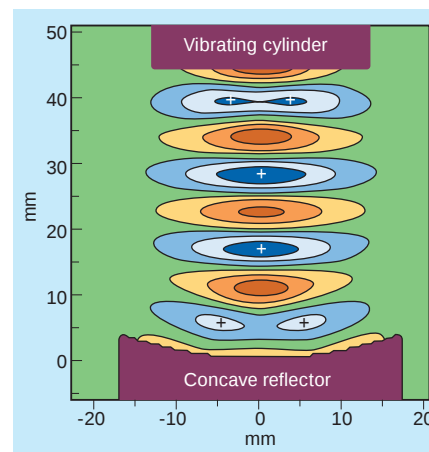


Figure 2 **Contour lines of the acoustic potential computed by Xie and Wei³. Tiny objects can be levitated in the four potential wells (or minima) denoted by crosses. The middle two wells are located on the central symmetry axis, and the top and bottom wells are ring shaped. The separation of the cylinder (top) and concave reflector (bottom) in the resonance mode is approximately two wavelengths, which is 20.3 mm for sound with a frequency of 16.7 kHz.**

tive Bernoulli pressure, which is proportional to the velocity squared. (Negative Bernoulli pressure provides the lifting force on an airplane wing.) But the acoustic force is a nonlinear effect; with a linear approximation of the model, any sound wave oscillates symmetrically over time, so all time averages would be zero and there would be no radiation force.

The acoustic potential calculated by Xie and Wei³ is shown as a contour plot in Fig. 2. This configuration corresponds to the resonant state in which the distance between emitter and reflector is approximately two sound wavelengths, which at 16.7 kHz is 20.3 mm for air at room temperature. The minima of the acoustic potential are marked by six crosses. As expected, these minima occur at four heights, close to the maxima of the velocity amplitudes, and their vertical spacing is approximately half a wavelength. If the sound intensity is large enough, each of these minima (or potential wells) can trap and levitate a small sample when the maximum vertical gradient of the potential (slope of the potential well) exceeds the gravitational force.

Xie and Wei³ show that the resonance frequencies and positions of stable levitation given by their model are in good agreement with the observed locations of the levitated spheres. The contours of the potential in Fig. 2 show that the two middle potential wells sit on the vertical axis of symmetry, but the wells next to the emitter and reflector form larger rings around the symmetry axis. This theoretical finding explains why the two samples at the top and bottom in Fig. 1b are located slightly off the central axis.

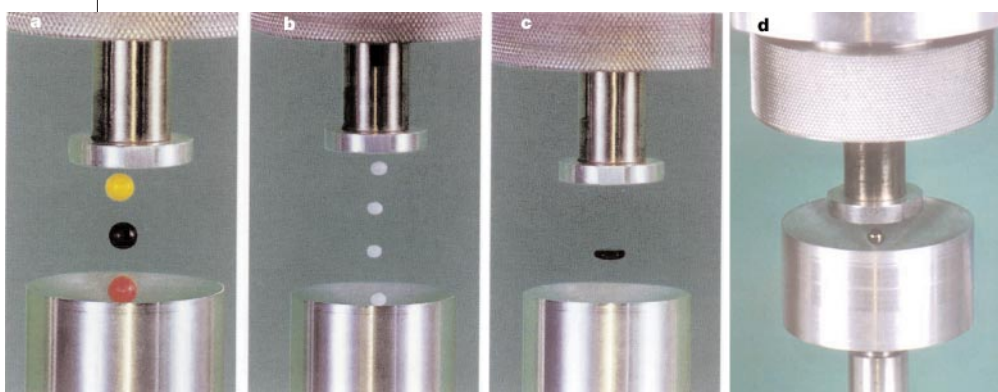


Figure 1 **Acoustic levitation by ultrasound. In Xie and Wei's latest experiment³ the force produced by a standing wave of 16.7-kHz ultrasound, which forms between a vibrating cylinder (top) and a concave reflector (bottom), is enough to levitate small objects. The freely floating objects are: a, three polymer spheres; b, four liquid crystal samples; c, a water drop deformed to a pancake shape by the uneven acoustic pressure; and d, a heavy tungsten sphere levitated in the lowest resonance mode, which has one potential minimum.**

Apart from acoustic levitation, there are several other types of contact-free levitation⁹: aerodynamic levitation (by a fluid jet), optical levitation (by a laser beam), electrical levitation (in a quadrupolar alternating electric field), radio-frequency levitation (by the eddy currents induced in a conducting sample by a conical coil), magnetic levitation (in the strong field of an electromagnet, which famously levitated a young frog^{10,11}), and superconducting levitation (by combining superconductors and permanent magnets). In all these types of free flotation, the main aim is to achieve good vertical and horizontal stability, otherwise a small perturbation may cause the levitating sample to fall.

Acoustic levitation has the advantage that it is simple and can levitate both non-magnetic and non-conducting materials. Experiments in microgravity in which metals and alloys are heated and cooled, without touching any container walls, have produced

new materials that cannot be formed by normal cooling, such as metallic glass and new superconductors. The improvements of the single-axis acoustic levitator achieved by Xie and Wei³ — increasing the levitation force with greater stability and predictability — may offer researchers the same opportunities as experiments in microgravity, but at a fraction of the cost. ■

E. H. Brandt is at the Max-Planck-Institute for Metals Research, D-70569 Stuttgart, Germany. e-mail: ehb@mf.mpg.de

1. Trinh, E. H. *Rev. Sci. Instrum.* **56**, 2059–2065 (1985).
2. Gammel, P. M., Cronquist, A. P. & Wang, T. G. *J. Acoust. Soc. Am.* **83**, 496–501 (1988).
3. Xie, W. J. & Wei, B. *Appl. Phys. Lett.* **79**, 881–883 (2001).
4. Bauerecker, S. & Neidhart, B. *J. Chem. Phys.* **109**, 3709–3712 (1998).
5. Weber, J. K. R. *et al. Rev. Sci. Instrum.* **65**, 456–465 (1994).
6. Xie, W. J. & Wei, B. *Chin. Phys. Lett.* **18**, 68–70 (2001).
7. Gor'kov, L. P. *Sov. Phys. Dokl.* **6**, 773–775 (1962); transl. (with misprints) from *Dokl. Akad. Nauk SSSR* **140**, 88–91 (1961).
8. Barmatz, M. & Collas, P. *J. Acoust. Soc. Am.* **77**, 928–945 (1985).
9. Brandt, E. H. *Science* **24**, 349–355 (1989).
10. Berry, M. V. & Geim, A. K. *Eur. J. Phys.* **18**, 307–313 (1997).
11. Brandt, E. H. *Phys. World* **10**, 23–24 (1997).

Human genetics

To clot or not

Amanda J. Fosang and Peter J. Smith

In multicellular animals, there has to be a balance between the free flow and clotting of blood. One molecule involved is von Willebrand factor, and the enzyme that cuts it down to size is now unveiled.

Thrombotic thrombocytopenic purpura is a potentially fatal human disease. It results from the widespread, abnormal aggregation of platelets and a protein, called von Willebrand factor, in the small blood vessels of many organs, including the brain and kidneys. The accumulated clots of platelets and proteins obstruct blood flow and cause red blood cells to fragment, leading to serious neurological and renal malfunctioning as well as anaemia, fever and a low platelet count (thrombocytopenia). On page 488 of this issue, Levy and colleagues¹ describe how they tracked down the gene that is mutated in patients with the hereditary form of this disorder. Their work adds significantly to our understanding of the molecules involved in blood clotting.

Von Willebrand factor (VWF) is a large protein that circulates in the blood as multimers of varying sizes. One of its main functions is to mediate interactions between platelets, and between platelets and the walls of blood vessels. Both types of interaction are vital in maintaining the balance between bleeding and clotting.

Thrombotic thrombocytopenic purpura (TTP) was first described² in 1924. But it was only recently linked to an increase in the number of unusually large multimers of VWF in the blood, which suggests that these abnormal forms might contribute to

the development of TTP. Then came the discovery that the activity of an enzyme that cleaves these large multimers into smaller fragments is markedly reduced in the blood of patients with TTP³. In some forms of the disease, patients produce an antibody that blocks the enzyme⁴; in others, the activity is missing altogether³. These observations pointed to the malfunctioning of a VWF-cleaving 'proteinase' as the underlying cause of the disease. Levy *et al.*¹ have now identified the gene, *ADAMTS13*, that encodes this proteinase, and show that mutations in this gene lead to an inactive enzyme and to TTP.

The first clues to the identity of the VWF-cleaving proteinase came about five years ago when the enzyme was partially purified, revealing that it requires zinc ions for full activity and is inhibited by metal-ion chelators^{5,6}. This partially purified enzyme could cleave VWF between the amino acids tyrosine at position 842 and methionine at position 843, generating characteristic fragments. Levy *et al.*¹ followed this proteinase activity to track down the affected gene in families with TTP. They developed an assay to measure how much of the cleavage product was present in patients' blood plasma.

This assay was critical to the authors' success. It allowed them to distinguish between

unaffected people (100% enzyme activity), unaffected individuals who were carriers of the mutant gene (50–60% activity) and TTP patients (2–7% activity). It also gave them the opportunity to undertake a 'linkage' study to pinpoint the chromosomal location of the defective gene, and at the same time to identify mutations that cause TTP. Linkage analysis needs families (pedigrees) with several generations of affected and unaffected people for study, as well as a phenotypic marker — in this case the proteinase activity. Armed with good pedigrees and a good marker, Levy *et al.* embarked on a search for patterns of inherited genes in families with TTP, matching these patterns with enzyme levels until they homed in on a single gene, *ADAMTS13*, on chromosome 9. In a single, elegant study they identified the enzyme, the gene, and the mutations in that gene that cause TTP. Other investigators, using protein-purification methods, have confirmed that *ADAMTS13* is the enzyme that processes VWF^{7,8}.

ADAMTS13 is a newly discovered member of a relatively new family of enzymes — the first *ADAMTS* protein was found only four years ago⁹. The name (which stands for 'a disintegrin-like and metalloproteinase with thrombospondin motifs') reflects the various modules, or motifs, that are found in these proteins (Fig. 1, overleaf). Each module is thought to contribute in some way to the protein's overall function. This must be true of *ADAMTS13* too, as the TTP-causing mutations found by Levy *et al.* are not clustered in one region, but are spread along more or less the entire length of the gene.

It is likely that one or more of the thrombospondin motifs in *ADAMTS13* helps it to dock onto VWF, while the metalloproteinase domain, which contains the active site, does the business of cleaving VWF. The role of the disintegrin-like domain is unknown. Disintegrins are molecules that bind integrin proteins on the surface of cells. Many snake venoms contain disintegrin domains that bind to integrins on platelets, leading to either haemorrhaging or clotting, depending on the venom. The disintegrin-like domain of the *ADAMTS* enzymes is about 40% identical to those of snake venoms. So far, no *ADAMTS* protein has been found to have a disintegrin-like function. But, given the relationship between platelets and VWF, it will be interesting to determine whether this region of *ADAMTS13* interacts with platelets.

We still have much to learn about the *ADAMTS* family. For example, deregulated expression of several *ADAMTS* proteins is implicated in disease (Fig. 1), but *ADAMTS2* and *ADAMTS13* are the only members of the family for which there is a clear association between an inherited disorder and a disturbed enzymatic process. Decreased levels of *ADAMTS2* prevent

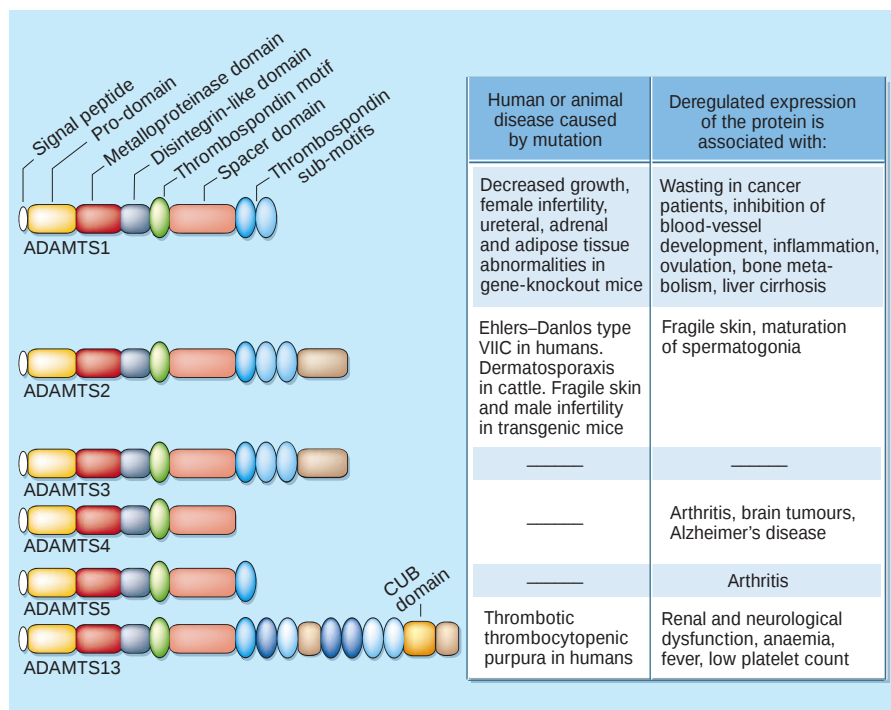


Figure 1 The ADAMTS family of proteins. All the ADAMTS enzymes described so far contain a 'signal peptide', which directs them to the cellular secretory apparatus. The pro-domain is cleaved off to produce the functional enzyme; the metalloproteinase domain contains the active site; and the function of the disintegrin-like domain is unknown. This family is distinguished from a similar group, the ADAM family, by the presence of thrombospondin type I motifs (regions first identified in the platelet protein thrombospondin). The spacer domain has both cysteine-rich regions and domains that lack cysteines. The thrombospondin submotifs show limited similarity to the thrombospondin type I repeat. Some of the enzymes also have unique domains that are unlike other protein modules. ADAMTS13 is the only member of the ADAMTS family to have a CUB domain — a feature of many extracellular proteins. ADAMTS 6, 7, 8 and 9 are orphan enzymes with no known substrates, and ADAMTS 10, 11 and 12 have been identified but not yet published.

the proper processing of type I collagen, leading to extremely fragile skin (Ehlers–Danlos syndrome type VIIC in humans and dermatosporaxis in cattle)¹⁰. Many of the other enzymes are 'orphans' with no known substrates.

Levy *et al.*¹ have shown that mutations in ADAMTS13 cause the inherited form of TTP. But the exact mechanism is not so clear, and the authors suggest several possibilities. The most likely scenario is that the persistence of high levels of large VWF multimers in the blood directly causes the disease. But it could be that the lack of fragmented products is important, or even that ADAMTS13 has other substrates in other tissues.

What is clear, though, is that new treatments for patients with TTP can now be developed. At the moment, therapy involves plasma exchange to replenish the active proteinase, remove multimeric VWF, and deplete any of the circulating ADAMTS13-specific antibody. Because the proteinase activity has a long half-life, of two to three days, this is an effective treatment¹¹. But now specific enzyme-replacement therapy becomes a viable alternative, especially given that as little as 5% of normal enzymatic activity may be sufficient to degrade large

VWF multimers¹². Finally, because the exact location of the ADAMTS13 gene within the human genome has been mapped, gene therapy — along the lines being explored for treating haemophiliacs — may be a realistic option in the future for patients with inherited TTP.

Amanda J. Fosang is in the Department of Paediatrics, University of Melbourne and Murdoch Childrens Research Institute, Royal Children's Hospital, Parkville 3052, Australia.

e-mail: fosang@cryptic.rch.unimelb.edu.au

Peter J. Smith is in the Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Parkville 3052, Australia.

e-mail: smithp@cryptic.rch.unimelb.edu.au

- Levy, G. G. *et al.* *Nature* **413**, 488–494 (2001).
- Moschowitz, E. *Proc. NY Pathol. Soc.* **24**, 21–24 (1924).
- Furlan, M. *et al.* *New Engl. J. Med.* **339**, 1578–1584 (1998).
- Tsai, H. M. & Lian, E. C. *New Engl. J. Med.* **339**, 1585–1594 (1998).
- Tsai, H. M. *Blood* **87**, 4235–4244 (1996).
- Furlan, M., Robles, R. & Lammie, B. *Blood* **87**, 4223–4234 (1996).
- Gerritsen, H. E., Robles, R., Lammie, B. & Furlan, M. *Blood* **98**, 1654–1661 (2001).
- Fujikawa, K., Suzuki, H., McMullen, B. & Chung, D. *Blood* **98**, 1662–1666 (2001).
- Kuno, K. *et al.* *J. Biol. Chem.* **272**, 556–562 (1997).
- Colige, A. *et al.* *Am. J. Hum. Genet.* **65**, 308–317 (1999).
- Furlan, M., Robles, R., Morselli, B., Sandoz, P. & Lammie, B. *Thromb. Haemost.* **81**, 8–13 (1999).
- Barbot, J. *et al.* *Br. J. Haematol.* **113**, 649–651 (2001).

Daedalus

Learning from melting

The most informative chemical test of all is simple heating; just warming the sample up and seeing what happens. Some substances, such as naphthalene, melt cleanly to a clear liquid; others, such as sucrose, give a hopeless mess confusing a change of phase with molecular alteration. Daedalus reckons that a shrewd chemist learns more about a material from a melting test than any textbook table tells him.

In 'melting with decomposition' the lattice melts first. The molecules, deprived of lattice support and about to decompose soon anyway, start to break up. The result is a discoloured melt, which, if allowed to cool at that point, gives a coloured solid with decomposed molecules in it. In 'decomposition with melting' the molecules decompose first. The lattice finds itself with nothing to order and falls apart, giving a deeply discoloured liquid that does not resolidify on cooling. This rich mixture melts at a much lower temperature, controlled by the 'lowered mixed-melting-point' principle. Yet textbooks confuse the two effects.

All this matters most, says Daedalus, in the industrial melting of impure mixtures, in particular in the manufacture of carbon fibres. These are made by stretching and heating a chosen pitch or polyacrylonitrile. Daedalus reckons that the ultimate performance of such fibres depends on the amount of polymeric carbon nanotube in them. Such tubes are much stiffer than steel, if not quite as strong for their weight.

If the lattice falls apart first, a useless mass of small molecules results. But if the molecules decompose first, there is a good chance that their product will be carbon nanotube, to be ordered by the surviving, stretching lattice. Polyacrylonitrile, sadly, has to eliminate ammonia to have any chance of forming carbon nanotubes; a tricky reaction. So DREADCO chemists are exploring other carbon-fibre reactions. They are heating polymers such as polyacrylamide and polyacrylic acid derivatives, hoping to find a reaction that gives more carbon nanotube and at a lower temperature. The stretching, orienting lattice could be an entirely different polymer; it might even be good old pitch. If they succeed, stronger carbon composites should soon result. This will lead to better fighter planes and fishing rods, more reliable cases and golf clubs, and stronger cars and computer casings. Even if they fail, Daedalus's chemists will still observe melting points with a clearer insight than current textbooks.

David Jones

Snapping shrimp make flashing bubbles

The cavitation bubbles created by shrimp in stunning their prey have some surprising properties.

Snapping shrimp produce a loud crackling noise^{1,2} that is intense enough to disturb underwater communication. This sound originates from the violent collapse of a large cavitation bubble generated under the tensile forces of a high-velocity water jet formed when the shrimp's snapper-claw snaps shut³ (Fig. 1). Here we show that a short, intense flash of light is emitted as the bubble collapses, indicating that extreme pressures and temperatures of at least 5,000 K (ref. 4) must exist inside the bubble at the point of collapse. We have dubbed this phenomenon 'shrimpluminescence' — the first observation, to our knowledge, of this mode of light production in any animal — because of its apparent similarity to sonoluminescence^{5,6}, the light emission from a bubble periodically driven by ultrasound.

To investigate this light-emission process, we positioned a snapping shrimp (*Alpheus heterochaelis*) in a seawater aquarium maintained at 20 °C and evoked a snap by gently stroking its snapper claw with a paintbrush. The sound pressure was recorded using a hydrophone in close proximity (1–3 cm) to the imploding bubble. The light emitted during the collapse was detected using a calibrated photodetector; all experiments were carried out in complete darkness.

The time course of the process is shown in Fig. 2a, b. The sound-pressure curve (Fig. 2a) consists of a small precursor signal, followed by an intense pressure peak at $t = 0$ that results from the collapse of the cavitation bubble (Fig. 1). A flash of light coincides with this pressure peak (Fig. 2b). A second pressure peak, originating from the collapse of a cloud of bubble fragments in the wake of the water jet some 300 microseconds after the main pressure peak, coincides with a second, dimmer light flash.

The flash duration is extremely short — shorter than could be temporally resolved in our experiment (limit, 10 nanoseconds; Fig. 2c). In a range of 100 nanoseconds around, but mainly before, the main light peak, very dim flashes also occur, which may be due to emission of light from the tiny bubble fragments (Fig. 1). Such multiple light pulses have also been observed for cavitating laser-induced bubbles^{7,8} and are associated with the non-sphericity of the collapse.

The total number of photons emitted from the hot bubble interior is up to 5×10^4 , which is one to two orders of magnitude less

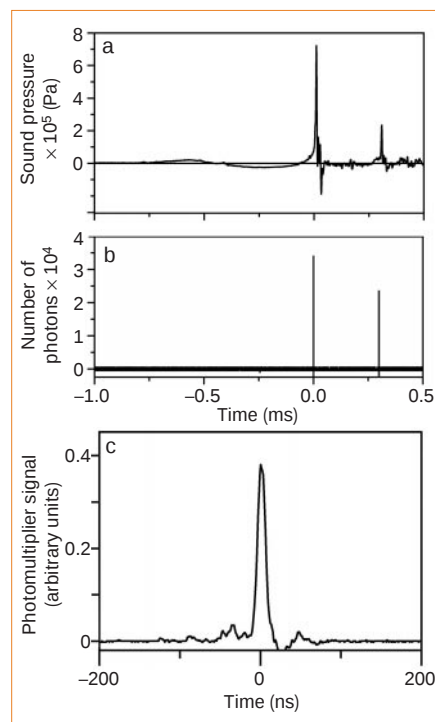


Figure 2 Sound and light from a snapping shrimp (*Alpheus heterochaelis*). **a**, Hydrophone signal and **b**, light emission from the snap of a shrimp. The principal light-emission event coincides with the bubble collapse at $t = 0$. A second light flash coincides with the subsequent collapse of a cloud of bubbles, which is formed upon collapse of the principal bubble. **c**, Expanded view of the photomultiplier signal, showing the short pulse duration of 'shrimpluminescence'. Within a range of 100 nanoseconds around the main peak, dimmer flashes of light are also evident; these may be due to emission of light from small bubble fragments formed during the violent collapse of the asymmetric cavitation bubble. Data analyses, R. M. Nelissen.

than the sonoluminescence typical of a single collapsing bubble. Shrimpluminescence cannot therefore be detected with the naked eye. Moreover, unlike sonoluminescence, the phenomenon is not repeated at a regular frequency, so there is no integration on the retina. A correlation of light intensity with sound intensity was not found in our set of 36 experiments with two different shrimp. We ascribe this to the uncontrolled way in which the asymmetric cavitation bubbles are created and then collapse.

The light emission associated with bubble collapse may not be biologically significant, but may instead represent a by-product of collapse, the shock wave from which is used to stun or even kill prey. But light emission is

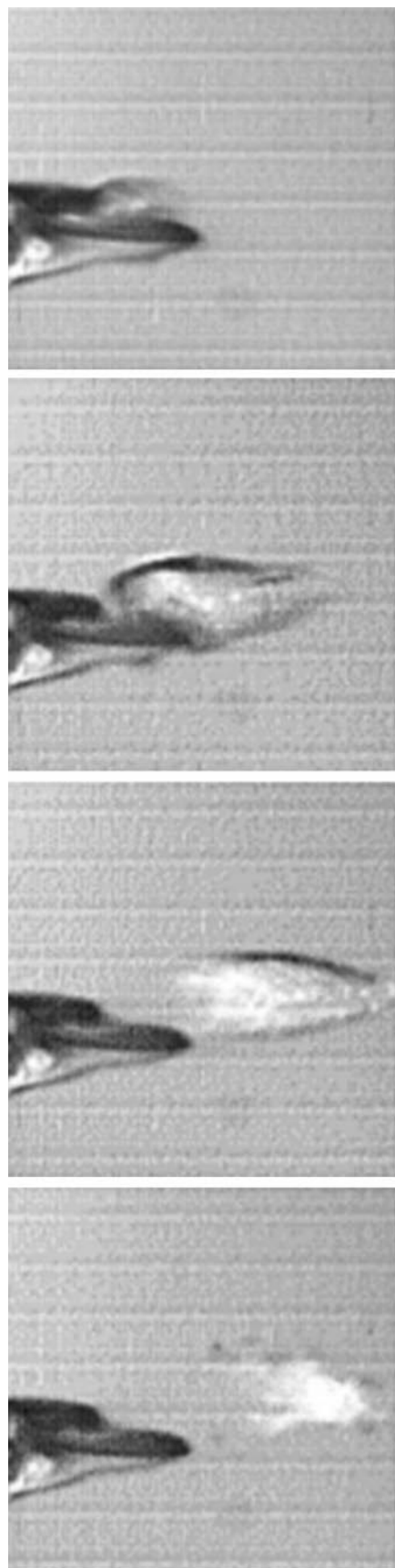


Figure 1 Side view of the tip of the snapper claw (left), showing the growth and subsequent collapse of the cavitation bubble, which produces a light flash and snapping sound. The light flash (not visible here) is indicative of the high temperature and pressure in the bubble interior at the point of collapse.

nonetheless indicative of the extreme conditions inside the bubble at collapse and therefore demonstrates the violence of the event.

Detlef Lohse*, Barbara Schmitz†, Michel Versluis*

*Faculty of Applied Physics and J. M. Burgers Research Centre for Fluid Dynamics, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands

e-mail: lohse@tn.utwente.nl

†Lehrstuhl für Zoologie, Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany

1. Everest, F. A., Young, R. W. & Johnson, M. W. *J. Acoust. Soc. Am.* **20**, 137–142 (1948).
2. Schmitz, B. in *Physiology of the Crustacean Nervous System* (ed. Wiese, K.) 521–533 (Springer, Berlin, 2001).
3. Versluis, M., Schmitz, B., von der Heydt, A. & Lohse, D. *Science* **289**, 2114–2117 (2000).
4. McNamara, W. B., Didenko, Y. T. & Suslick, K. S. *Nature* **401**, 772–775 (1999).
5. Brennen, C. E. *Cavitation and Bubble Dynamics* (Oxford Univ. Press, New York, 1995).
6. Crum, L. A. *Phys. Today* **47**, 22–29 (1994).
7. Ohl, C. D. *Zur Dynamik und Lumineszenz von Kavitationsblasen*. Thesis, Georg August Univ., Göttingen (1999).
8. Baghdassarian, O., Tabbert, B. & Williams, G. A. *Phys. Rev. Lett.* **83**, 2437–2440 (1999).

Carbon emissions

The economic benefits of the Kyoto Protocol

The third Conference of the Parties in Kyoto set the target of reducing greenhouse-gas emissions by an average of 5.3% with respect to 1990 values by 2008–2012. One of the main objections to the protocol's ratification is that compliance would pose an unbearable economic burden on the countries involved¹. But we show here that this is not the case if costs apart from the direct costs of energy production are also considered. Costs are also incurred in rectifying damage to human health, material goods, agriculture and the environment related to greenhouse-gas emissions.

We have analysed alternative scenarios for electricity generation² (which contributes roughly one-third of total greenhouse-gas (GHG) emissions) in Italy in about 10 years' time (2010) by considering different technologies, including co-generation (combined production of electric and thermal energy). These technologies are based on oil, solid fossil fuels such as coal, gas (methane, for example) and renewable sources such as hydropower, wind, photovoltaic or biomass.

We accounted both for direct, annual industrial costs and for social and environmental costs. These latter are 'externalities' to the energy producers³ and are not included in companies' balance sheets, but need to be considered in national balances. These external costs can be classified as local (derived from direct damage to the country in question — for example, as a result of increased air and water pollution and land deterioration) and global (derived from damage to the entire biosphere by GHG emissions).

We improved recent estimates⁴ of production costs using experience curves⁵ (diminishing costs due to technological innovation) and assessed external costs through a comparative analysis of the three main studies conducted in Europe (Project ExternE, probably the best and most comprehensive analysis)⁶ and the United States^{7,8}. The uncertainty of local external

costs is reasonably small, whereas that associated with global costs is much larger⁶ (4–140 euros per tonne of CO₂). We therefore carried out a sensitivity analysis of situations with global costs ranging from zero to 250 euros per tonne CO₂.

We have examined three situations (see supplementary information), each corresponding to a different problem of cost minimization and subject to two constraints: the estimated energy demand of Italy in 2010 (353 terawatt-hours) should be satisfied; and each energy-production technology should not exceed a maximum feasible quota (estimated on the basis of physical constraints, the current state of technology and its predicted progress²). The three problems of cost minimization correspond to: minimizing only the sum of production costs (business as usual, BAU); minimizing total social costs (MSC; that is, the sum of industrial and external costs of energy production); and minimizing only production costs, with the further constraint that the Kyoto Protocol must be satisfied (BAU + Kyoto). The target set for Italy is a 6.5% reduction in GHG emissions with respect to 1990 levels. Here, as we are considering only electricity generation, the BAU + Kyoto scenario does not require Italy as a whole to satisfy the Kyoto protocol, but merely constrains the electricity sector to cut its GHG emissions by 6.5%.

If we use the average global external cost calculated by the ExternE⁶ study (30 euros per tonne CO₂), we obtain the results shown in Fig. 1. The BAU situation implies much higher external costs than the other two, not only in terms of uncertain global costs, but also in the much more certain local costs. Both MSC (which happens to satisfy the Kyoto Protocol) and BAU + Kyoto require large quotas of gas, which has a high combustion efficiency and lower emission of pollutants such as NO_x and SO₂, which are responsible for most local external costs. MSC, however, uses far less coal. We also carried out a sensitivity analysis for MSC (see supplementary information) and found that the value of shares in the different technologies would not change markedly, but that increasing global external costs would cause a shift from gas towards

renewable sources and co-generation.

The estimated balance of BAU + Kyoto against BAU is a 17% reduction in GHG emissions, 1,829 million euros saved in reducing external costs (local, 1,068 million; global, 761 million), and a cost of 308 million euros in increased industrial outlay. However, if the cost of GHG reduction is smaller in the electricity sector than in those responsible for the remaining two-thirds of emissions, the industrial costs of compliance with the Kyoto Protocol for Italy might be more than three times those shown here.

The economic costs of global warming cannot be accurately estimated and may never be⁹. However, even accounting only for local external costs, together with production costs, to identify energy strategies, compliance with the Kyoto Protocol would imply lower, not higher, overall costs. Also, reducing local and regional air pollution is sufficient reason to cut coal combustion, which is the single worst GHG source from a global perspective.

Whatever decision-makers believe about climate change, they should make changes to their power-generation strategies now. Active energy-conservation policies, which are not considered here, might further increase social benefits without harming economies¹⁰. We conclude that any delay in reducing GHG

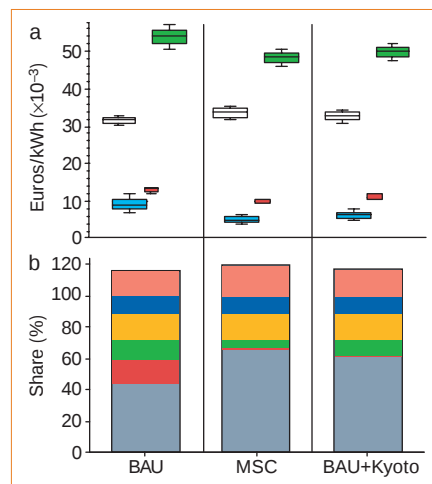


Figure 1 Annual costs of one kilowatt-hour of electricity in Italy and the shares of technologies for three different cost situations (see text). Box plots were obtained by considering the range of estimated costs for each technology (see supplementary information). Values in **a** represent: white, industrial costs; blue, local external costs; red, global external costs; green, total costs. Values in **b** represent: grey, gas; red, liquid fossil fuel; green, solid fossil fuel; yellow, renewable sources; blue, imported (nuclear); pink, co-generation. Of Italy's energy requirements, 11.3% is produced abroad (mainly nuclear energy imported from France and Switzerland). Shares of the technologies total more than 100%, because part of the energy is co-generated, which reduces external costs. With respect to the 'business as usual' (BAU) situation, compliance with the Kyoto Protocol implies a 3.1% increase in industrial costs but a 35% reduction in external costs — a net saving of 1.5 billion euros per year (8.8% of the total BAU cost). The saving in the 'minimize total social costs' (MSC) situation is 2 billion euros, or 11.7% of the total BAU cost.

emissions is not justified.

Giulio A. De Leo*, Luca Rizzi†, Andrea Caizzi†, Marino Gatto§

*Dipartimento di Scienze Ambientali, Università degli Studi di Parma, Parco Area delle Scienze, Parma 43100, Italy
e-mail:deleo@dsa.unipr.it

†Centro Elettrotecnico Sperimentale Italiano, Business Unit Ambiente, Via Reggio Emilia 39, Segrate 20092, Italy

§Centro di Ingegneria Biomedica, Consiglio Nazionale delle Ricerche, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano 20133, Italy

1. Remarks by President G. W. Bush on Global Climate Change (Office of the Press Secretary, The White House, 11 June 2001).
2. De Leo, G. A., Rizzi, L. & Caizzi, A. in *Proc. 29th Int. Conf. Automation and Decision Making* 247–256 (FAST, Milan, 2000).
3. Callan, S. J. & Thomas, J. M. *Environmental Economics and Management: Theory, Policy and Applications* (Irwin, Chicago, 1996).
4. De Paoli, L. & Lorenzoni, A. *Economia e politica delle fonti rinnovabili e della cogenerazione* (Franco Angeli, Milan, 1999).
5. IEA/OECD Experience Curves for Energy Technology Policy (IEA, Paris, 2000) (<http://www.iea.org/public/studies/curves.htm>).
6. European Commission ExternE: *Externalities of Energy Vols 1–6* (EC, Brussels, 1995) (<http://extern.e.jrc.es/index.html>).
7. Rowe, R. D. et al. *ESEERCO, New York State Environmental Externalities Cost Study Vols 1, 2* (Oceana, New York, 1995).
8. Russell, L. (ed.) *External Costs and Benefits of Fuel Cycles* (Oak Ridge National Laboratory and Resources for the Future, Oak Ridge, Tennessee, 1994).
9. Gatto, M. & De Leo, G. A. *Bioscience* **50**, 347–355 (2000).
10. Interlaboratory Working Group *Scenarios for a Clean Energy Future* (Oak Ridge National Laboratory, Tennessee, and Lawrence Berkeley National Laboratory, California; OPRNL/CON-476 and LBNL-44029, 2000).

Supplementary information is available at <http://www.nature.com> or as paper copy from the London editorial office of *Nature*.

Palaeolithic paintings

Evolution of prehistoric cave art

Sophisticated examples of European palaeolithic parietal art can be seen in the caves of Altamira, Lascaux and Niaux near the Pyrenees, which date to the Magdalenian period (12,000–17,000 years ago), but paintings of comparable skill and complexity were created much earlier^{1,2}, some possibly more than 30,000 years ago³. We have derived new radiocarbon dates for the drawings that decorate the Chauvet cave in Vallon-Pont-d'Arc, Ardèche, France, which confirm that even 30,000 years ago Aurignacian artists, already known as accomplished carvers¹, could create masterpieces comparable to the best Magdalenian art⁴. Prehistorians, who have traditionally interpreted the evolution of prehistoric art as a steady progression from simple to more complex representations, may have to reconsider existing theories of the origins of art.

The chronology of European prehistoric cave paintings has been loosely based on the style of fauna depicted or on dated remains left by cave occupants, but has become more precise with radiocarbon dating of the char-

Figure 1 'Horse' panel from the Hillaire chamber of the Chauvet cave in Vallon-Pont-d'Arc, Ardèche, France, which shows a rhinoceros and was drawn more than 30,000 years ago.



coal pigments themselves. Accelerator mass spectrometry, which relies on the separation and counting of carbon isotopes, requires much less of this precious sample material than traditional ¹⁴C-dating techniques.

Uncalibrated radiocarbon ages in excess of 22,000 years (22 Kyr before present (BP)) have been calculated for paintings in several French caves, mostly on the tandem accelerator at Gif-sur-Yvette, France⁵. The charcoal of the painting itself can be dated⁶, either directly or in trace organic residues that have a temporal relationship to the paintings, for example as charcoal mingled with ochre pigments. Also datable are smudges left by torch-bearers which, if found on the calcite coating of a drawing, indicate a time before which the drawing must have been created.

Indirect evidence of extensive painting activity before the Solutrean period comes from radiocarbon dates for drawings at two French caves — a 26.9-Kyr-old bone chip was extracted from a fissure crossing a stencilled hand at Gargas⁷, and three burnt bones, mixed with red and yellow ochre at the base of several designs in the Grande Grotte at Arcy-sur-Cure⁸, are aged at 26–28 Kyr; torch smears marring the red frieze date to about 27 Kyr BP⁹.

At four other French caves, charcoal from the drawings themselves has been dated. During investigations at Cougnac, dates between 23 and 25 Kyr BP were obtained for two giant deer¹⁰, and at nearby Pech Merle, the right-facing spotted horse is dated to 24.7 Kyr BP¹. In Cosquer, we dated 13 drawings^{5,11} at 27–28 Kyr BP for two stencilled hands, a bison and an oval sign — all other drawings but two were 18–20 Kyr old (cave-floor charcoal fits within the same two periods⁵).

In the Chauvet caves, which consist of several chambers, we derived radiocarbon dates of between 29.7 and 32.4 Kyr BP for charcoal (0.27–1.40 mg carbon) from animals painted in the Salle du Fond and in the 'horse' panel (Fig. 1) of the Hillaire chamber⁴. Two torch rubbings, one from the same panel and another from the Cierge chamber, were about 27 Kyr old, a

reasonable age considering that in one case the torch was scraped against a calcite-coated animal. We obtained an age of 31.4 Kyr for a giant deer at the entrance to the Megaceros gallery (see supplementary information). Charcoal was obtained from under a bear skull placed on a stone slab in the Crâne chamber and from the Megaceros gallery, which is carpeted with charcoal particles of various sizes as though it had been used as a charcoal factory⁴. Apart from two roughly 26-Kyr-old specimens, most of the charcoal was produced between 29 and 32 Kyr BP, suggesting that there may have been two significant episodes of human intrusion before the cave was sealed off by a rockfall.

This latest comprehensive dating confirms our earlier provisional assignment of the Chauvet cave art to the Aurignacian period³. Future discoveries will reveal the nature of paintings that predate those in the Chauvet cave.

H. Valladas*, J. Clottes†, J.-M. Geneste‡, M. A. Garcia§, M. Arnold¶, H. Cachier*, N. Tisnérat-Laborde*

*Laboratoire des Sciences du Climat et de l'Environnement, UMR CEA-CNRS 1572, 91198 Gif-sur-Yvette Cedex, France
e-mail: helene.valladas@lscce.cnrs-gif.fr

†11 Rue du Fourcat, 09000 Foix, France

‡Service Régional de l'Archéologie d'Aquitaine, DR AC d'Aquitaine, 54 Rue Magendie, 33074 Bordeaux, France

§Maison de l'Archéologie et de l'Ethnologie, 21 Allée de l'Université, 92023 Nanterre, France

¶UMS 2004 (CNRS-CEA), Tandétron Bâtiment 30, 91198 Gif-sur-Yvette, France

1. Lorchlanche, M. *Les Grottes Ornées de la Préhistoire* (Errance, Paris, 1995).
2. Clottes, J. & Courtin, J. *La Grotte Cosquer* (Seuil, Paris, 1994).
3. Clottes, J. et al. *C. R. Acad. Sci. Paris* **320**, 1133–1140 (1995).
4. Clottes, J. (ed.) *La Grotte Chauvet: L'art des Origines* (Seuil, Paris, 2001).
5. Valladas, H. et al. *Radiocarbon* (in the press).
6. Valladas, H. et al. *Nature* **357**, 68–70 (1992).
7. Clottes, J. & Valladas, H. *Bull. Soc. Préhist. Fr.* **89**, 270–274 (1992).
8. Girard, M. et al. *Cahiers Archéologiques de Bourgogne* **6**, 17–23 (1996).
9. Baffier, D. et al. *Int. Newslett. Rock Art* **28**, 1–3 (2001).
10. Valladas, H., Cachier, H. & Arnold, M. *Rock Art Res.* **7**, 18–19 (1990).
11. Clottes, J. et al. *Bull. Soc. Préhist. Fr.* **89**, 230–234 (1992).

Supplementary information is available at <http://www.nature.com> or as paper copy from the London editorial office of *Nature*.

Physiology

Cold current in thermoreceptive neurons

We sense the temperature of our skin and surroundings using specific thermoreceptors, which are sensitive to cold and warmth¹, but little is known about how these receptors transduce temperature into electrical activity. We have discovered an inward ionic current that is activated by moderate cooling in a small number of rat sensory neurons. This current has features that are found in intact cold receptors, including sensitization by menthol, adaptation upon sustained cooling, and modulation by calcium, and is likely to be important in cold sensing.

Cutaneous receptors are difficult to study, as they are small and inaccessible in the skin, so cultured dorsal root ganglion (DRG) neurons are widely used as a model system because they express membrane proteins that would normally be present at their receptor termini. We found previously that very few DRG neurons generate action potentials in response to cold², so we measured the intracellular calcium-ion concentration by imaging³ to preselect cold-responsive rat DRG neurons after 2–4 days in primary culture, applying thermal stimuli with a Peltier-based device⁴ from a base temperature of 32 °C. Of 643 DRG neurons, 45 (7%) responded to cooling to 20 °C by increasing their intracellular Ca²⁺ concentrations. This is consistent with the proportion of cold thermoreceptor afferents in the rat hindlimb^{5,6}.

In the cold-responsive DRG neurons, we measured membrane potential and ionic

currents using perforated-patch recording. We raised the temperature (for about 10 s) from 32 °C to 37 °C, and then applied a ramp from 37 °C to 18–20 °C over a 30-s period. From a resting potential of -53.3 ± 7.9 mV (s.d.; $n = 32$), the cooling ramp induced a depolarization of 7–49 mV (22.3 ± 11.2 mV) and high-frequency action potentials (Fig. 1a). Cooling at -80 mV elicited an inward current (Fig. 1b) with a threshold of 23–34 °C (28.7 ± 2.7 °C; $n = 27$) and a maximum amplitude of 40–350 pA.

This current was present in all cold-responsive neurons tested, and was absent from all of 16 unresponsive ones, showing that it is not a ubiquitous current in primary somatosensory neurons. By applying temperature steps, we found that the activation and deactivation of this current were as rapid as (and perhaps limited by) the time course of the thermal stimulus, which has a time constant of about 5 s (ref. 4). During sustained cold stimulation, the current adapted almost completely, with a time constant of 62–69 s ($n = 4$; Fig. 1), consistent with the slow adaptation of cold receptors *in vivo*⁷.

The cold-receptor stimulant (–)menthol^{8,9} substantially increased the amplitude of the cold-induced current and shifted its activation threshold towards higher temperatures, by 4.2 ± 2.0 °C at 10 μ M ($n = 16$) and by 7.6 ± 2.6 °C at 100 μ M ($n = 9$; Fig. 1b). A low extracellular Ca²⁺ concentration potentiated the current (1.5–2.8-fold at 0.1 mM; $n = 6$), whereas increasing the extracellular Ca²⁺ concentration to 10 mM reduced it slightly and reversed the sensitization induced by 10 μ M menthol ($n = 4$); intact cold receptors are similarly affected by changes in Ca²⁺ concentration^{9,10}. All of the effects due to

menthol and altered Ca²⁺ were reversible.

The reversal potential of the cold- and menthol-induced current obtained by subtraction during voltage ramps was $+13.3 \pm 5.2$ mV ($n = 6$), indicating that it is probably a mixed-cation current. We cannot assign it to one of the known families of cation channels involved in somatosensory transduction, as amiloride (100 μ M), which blocks acid-sensitive and mechanosensitive channels of the degenerin family, had no effect ($n = 6$), whereas ruthenium red (10 μ M), a blocker of the heat-sensitive VR-1 and VRL-1 channels, reversibly increased the current ($n = 6$).

Our observations suggest that the mode of action of menthol on cold receptors should be reconsidered. An early model indicated that menthol stimulates cold receptors by blocking voltage-dependent Ca²⁺ channels, leading to a reduction in intracellular Ca²⁺ and inhibition of Ca²⁺-dependent K⁺ channels⁹. However, we and others¹¹ have since discovered that menthol stimulates entry of Ca²⁺ and increases intracellular Ca²⁺ concentration in cold-sensitive neurons; the stimulation of cold receptors by menthol can be explained more simply by sensitization of the cold-induced inward Ca²⁺ current.

To our knowledge, this is the first description of an ionic current that is activated by cooling. The properties of this current can account for several features of cold-receptor function: it is activated over the temperature range in which mammalian cold receptors are most sensitive¹; its rate of adaptation is similar to that observed *in vivo*⁷; and its potentiation by menthol and modulation by calcium are similar to the response of intact cold receptors^{8–10}. We propose that this cold-activated current is the principal determinant of cold-receptor activity, and that inhibition of a background K⁺ current² and of the electrogenic Na⁺/K⁺-ATPase¹² are of secondary importance.

Gordon Reid, Maria-Luiza Flonta

Department of Animal Physiology and Biophysics,
Faculty of Biology, University of Bucharest, Splaiul
Independenței 91–95, 76201 Bucharest, Romania
e-mail: gordon@biologie.kappa.ro

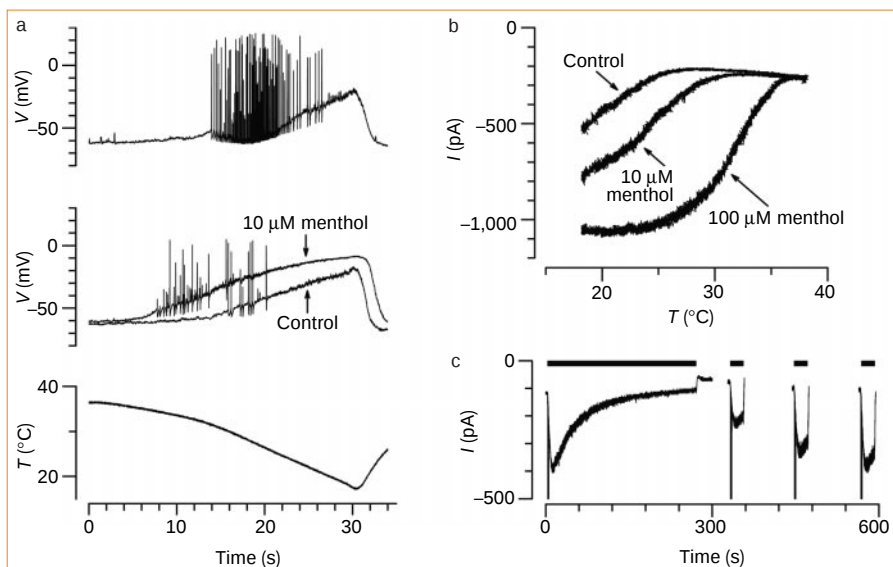


Figure 1 Cold sensitivity and cold-induced inward current in primary sensory neurons. **a**, Top, depolarization and action potentials during cooling in a rat dorsal root ganglion neuron; middle, sensitization by 10 μ M (–)menthol in the same neuron (10 μ M bupivacaine was added to reduce action-potential frequency); bottom, thermal stimulus. **b**, Current–temperature relationship of cold-induced current during ramps similar to those in **a**, and the sensitization induced by 10 μ M and 100 μ M (–)menthol. **c**, Adaptation of the cold-induced current during prolonged cooling to 15 °C (horizontal bars) and recovery at 32 °C; action potentials in cell processes are apparent.

- Hensel, H. *Thermoreception and Temperature Regulation* (Academic, London, 1981).
- Reid, G. & Flonta, M.-L. *Neurosci. Lett.* **297**, 171–174 (2001).
- Suto, K. & Gotoh, H. *Neuroscience* **92**, 1131–1135 (1999).
- Reid, G., Amuzescu, B., Zech, E. & Flonta, M.-L. *J. Neurosci. Methods* **111**, 1–8 (2001).
- Gee, M. D., Lynn, B., Basile, S., Pierau, F. K. & Cotterell, B. *Neuroscience* **90**, 509–518 (1999).
- Leem, J. W., Willis, W. D. & Chung, J. M. *J. Neurophysiol.* **69**, 1684–1699 (1993).
- Kenshalo, D. R. & Duclaux, R. J. *Neurophysiol.* **40**, 319–332 (1977).
- Hensel, H. & Zotterman, Y. *Acta Physiol. Scand.* **24**, 27–34 (1951).
- Schäfer, K., Braun, H. & Isenberger, C. J. *Gen. Physiol.* **88**, 757–776 (1986).
- Hensel, H. & Schäfer, K. *Pflügers Arch.* **352**, 87–90 (1974).
- Okazawa, M., Terauchi, T., Shiraki, T., Matsumura, K. & Kobayashi, S. *NeuroReport* **11**, 2151–2155 (2000).
- Pierau, F., Torrey, P. & Carpenter, D. *Brain Res.* **73**, 156–160 (1974).

Warm tropical sea surface temperatures in the Late Cretaceous and Eocene epochs

Paul N. Pearson*, Peter W. Ditchfield*, Joyce Singano†, Katherine G. Harcourt-Brown*, Christopher J. Nicholas‡, Richard K. Olsson§, Nicholas J. Shackleton|| & Mike A. Hall||

* Department of Earth Sciences, University of Bristol, Queen's Road, Bristol BS8 1RJ, UK

† Tanzania Petroleum Development Corporation, PO Box 2774, Dar-es-Salaam, Tanzania

‡ Department of Geology, Trinity College, Dublin 2, Ireland

§ Department of Geological Sciences, Rutgers University, Piscataway, New Jersey 08855-1179, USA

|| Godwin Laboratory, Department of Earth Sciences, University of Cambridge, Free School Lane, Cambridge CB2 3SA, UK

Climate models with increased levels of carbon dioxide predict that global warming causes heating in the tropics, but investigations of ancient climates based on palaeodata have generally indicated cool tropical temperatures during supposed greenhouse episodes. For example, in the Late Cretaceous and Eocene epochs there is abundant geological evidence for warm, mostly ice-free poles, but tropical sea surface temperatures are generally estimated to be only 15–23 °C, based on oxygen isotope palaeothermometry of surface-dwelling planktonic foraminifer shells. Here we question the validity of most such data on the grounds of poor preservation and diagenetic alteration. We present new data from exceptionally well preserved foraminifer shells extracted from impermeable clay-rich sediments, which indicate that for the intervals studied, tropical sea surface temperatures were at least 28–32 °C. These warm temperatures are more in line with our understanding of the geographical distributions of temperature-sensitive fossil organisms and the results of climate models with increased CO₂ levels.

Since the discovery that the oxygen isotope ratio ($\delta^{18}\text{O}$) of calcite is dependent on the temperature of the water from which it is precipitated^{1,2}, many studies have used the technique as a thermometer for the ancient oceans by analysing the shells of fossil planktonic foraminifera and other organisms. As has been widely debated^{3,4}, in order to produce quantitative temperature estimates it is necessary to consider factors such as isotope fractionation by the organism, the oxygen isotopic composition of ocean water, global ice volume, the evaporative, precipitation and mixing history of the local water mass, and variations of pH and alkalinity. It is also necessary to assume that the original shell $\delta^{18}\text{O}$ does not change over time, but this is difficult to assess and strict preservational criteria have seldom been applied. The quality of preservation generally declines with geological age and burial depth in the cored pelagic oozes and chalks that are the principal resource for palaeoceanographic studies^{5,6}. Most workers have accepted estimates of sea surface temperature (SST) from samples in which clear species differences are preserved between surface, deeper-dwelling and benthic taxa, and, for time-ordered data sets, in which reasonable values—and trends and events that can be correlated—are observed^{3,7}. Unfortunately, this does not guarantee unaltered isotopic compositions because the possibility of a component of diagenetic calcite cannot be excluded.

After the first phase of ocean drilling, it became evident that the oxygen isotope technique predicted surprisingly cool Eocene tropical SSTs from planktonic foraminifera, while at the same time indicating relatively warm SSTs at higher latitudes and warm deep-water temperatures (based on analysis of benthic foraminifera)^{7–9}. This led to the suggestion that more efficient poleward heat transport might have occurred in the past, perhaps due to a fundamentally different mode of ocean circulation^{7,10,11}. In this way it might even be possible to warm the poles sufficiently to prevent ice build-up without invoking a higher global heat budget. However, it has been calculated that an excessive heat flux would be required to flatten latitudinal gradients sufficiently¹², and all climate

models to date—using modern or ancient reconstructed ocean basin geometries^{13–17}—have failed to simulate such a climate regime. Although ocean circulation no doubt has an important influence on global climate through albedo-related feedbacks, it seems evident that an increased solar constant or enhanced greenhouse-gas concentrations are required to heat the poles to the required degree, and tropical warming has always attended such models.

The mismatch between data and models has become more acute as further tropical deep-sea carbonates have been drilled and analysed for $\delta^{18}\text{O}$ (refs 18, 19). A very similar pattern has also emerged for the Late Cretaceous epoch^{20,21} and, in more general terms, for oxygen isotope measurements of carbonate sediments from warm climates spanning the past 500 Myr (refs 22, 23). This has led some authors to question the primary role of carbon dioxide in determining global temperatures²², while others have maintained that the $\delta^{18}\text{O}$ measurements may be biased toward positive values because of early diagenesis in the cold sea-floor environment^{5,6,24,25}.

A recent detailed evaluation of the diagenetic problem concluded that overgrowth and infilling of foraminifer shells by calcite could reduce apparent temperatures by a maximum of 5 °C and by probably only 1–2 °C for a typical sample³, which is far short of the values necessary to resolve the discrepancy. Although microscopic recrystallization is also widely recognized as a possible problem, its effect is more difficult to quantify³. The most persuasive argument in support of the cool temperature interpretation of planktonic foraminiferal $\delta^{18}\text{O}$ follows from the expectation that the carbon isotopic composition ($\delta^{13}\text{C}$) of diagenetic calcite should be around +3‰ (a typical bulk sediment value), which is within the range of reported inter-species values for planktonic foraminifera. Hence, although additional diagenetic calcite might plausibly displace $\delta^{18}\text{O}$ measurements to more positive values while retaining most of the inter-species differentials, the $\delta^{13}\text{C}$ differentials would be expected to rapidly collapse if such overprinting occurred³.

A representative example of a recrystallized Middle Eocene

multi-species data array²⁶ is shown in Fig. 1, along with an interpretative model for the depth habitats of the respective species. Significant oxygen and carbon isotope differentials between the species are observed, which are explained by their different ecologies. These differentials are consistent with many other sites in which the same species have been analysed. Furthermore, species-specific relationships between carbon isotope compositions and size are also observed in surface-dwelling taxa in this and similar samples²⁶. Data such as these have led most authors to conclude that substantial diagenetic overprinting of the isotopic ratios is unlikely, even when the microstructural preservation is acknowledged as imperfect^{3,26}.

However, while the vast majority of $\delta^{18}\text{O}$ estimates have supported the cool tropics view, occasional studies of exceptionally well preserved mollusc shells^{25,27}, early diagenetic cements²⁵, and phosphates²⁸ have indicated considerably warmer temperatures approaching 30 °C. In addition, recent studies^{29,30} of exceptionally well preserved planktonic foraminifera have indicated similar warm temperatures in the Middle Cretaceous. These results have prompted us to collect rare, exceptionally well preserved assemblages from the Late Cretaceous and Eocene of Tanzania and elsewhere, to further test the validity of planktonic foraminiferal SST estimates for these periods.

Shell construction and diagenesis in foraminifera

The primary shell wall in planktonic foraminifera is constructed from aggregates of morphologically discontinuous microgranules or plaques of sub-micrometre size that are originally precipitated intra-cellularly close to the site of chamber formation³¹. The microgranules are then fused or bound together by further calcite growth creating a smooth surface, often with finely sculpted features such as pores, hexagonal pore-pits, surface pustules and radiating spines (Fig. 2a–f). Planktonic shells appear to have substantial microscopic porosity between the microgranules, whereas benthic foraminifera tend to have denser, more robust shells. The relatively low-density shell structures of planktonic foraminifera may be an adaptation for maintaining buoyancy.

High-resolution scanning electron microscope (SEM) images³¹ show that planktonic foraminifer microgranules have irregular shapes and consequently high surface areas, making them prone

to diagenetic recrystallization. In this way, a shell may become thoroughly recrystallized on a micrometre scale without obliterating structures such as wall pores, internal shell layering features and surface ornament. Our observations of many Cretaceous and Palaeogene samples suggests that this kind of diagenesis is ubiquitous in the pelagic oozes and chalks that are commonly used in palaeoceanographic studies, even when preservation has been described as good or excellent (Fig. 2g–i). We emphasize that this process is distinct from a more obvious mode of alteration that involves the precipitation of euhedral inorganic calcite crystals from solution on the inner and outer surfaces of the shells, sometimes resulting in substantial infillings or overgrowths. Such overgrowths are also very common in more deeply buried chalk and limestone samples, but have generally been avoided in previous isotopic work.

Pristine planktonic foraminifer shells of small modern non-encrusting species or juveniles are translucent and appear 'glassy' under the light microscope. This kind of preservation is rare in ancient material. 'Glassy' preservation does, however, occur in some impermeable clay-rich sediments that have escaped significant burial^{29,30}. The reason for this appears to be that clay-rich sediments have very low porosity and permeability, so that carbonate fossils are effectively sealed within the sediment and quickly equilibrate with their accompanying fluids.

Certain localities have been prized by micropalaeontologists for their exceptional preservation but have not previously been targeted in isotopic work. In particular, outcrops of Cretaceous and Palaeogene silty clays in southern coastal Tanzania have long been known to yield excellently preserved material from a wide range of ages^{32,33}. We re-collected these sites with a view to conducting new stable isotope studies and testing what has become known²⁰ as the 'cool tropic paradox'. We also analysed three other exceptionally well preserved Eocene assemblages from Mexico, Alabama and the Adriatic Sea, allowing us to examine other tropical to mid-latitude sites. Details of the localities, their geological setting and micropalaeontological characterization are available; see Supplementary Information.

Stable isotope results

We selected samples for analysis in which there is no significant overgrowth or infilling of diagenetic calcite, and in which

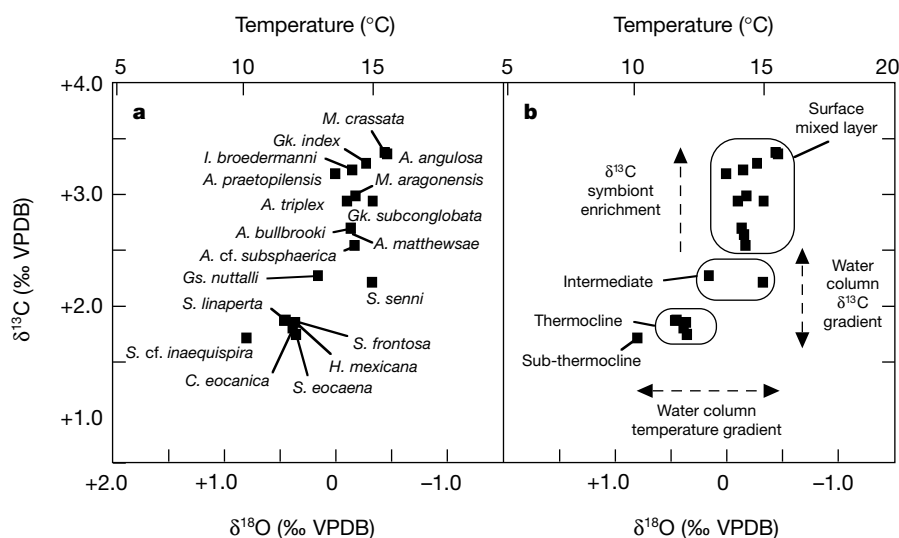


Figure 1 Multi-species stable isotope array from a typical recrystallized planktonic foraminifer assemblage from the Middle Eocene (DSHP Site 523, Angola basin, South Atlantic Ocean, modified from ref. 26). **a**, Cross-plot of mean oxygen (scale reversed) and carbon isotope ratios of each species. Note the retention of substantial carbon i isotope differentials despite diagenesis. **b**, Interpretative model for the respective depth-habitats

of the different species. Burial depth is 150–180 m. Absolute temperature estimates and $\delta^{13}\text{C}$ values are affected by diagenesis. $\delta^{18}\text{O} = [({}^{18}\text{O}/{}^{16}\text{O})_{\text{sample}} / ({}^{18}\text{O}/{}^{16}\text{O})_{\text{standard}}] - 1$, $\delta^{13}\text{C} = [({}^{13}\text{C}/{}^{12}\text{C})_{\text{sample}} / ({}^{13}\text{C}/{}^{12}\text{C})_{\text{standard}}] - 1$, VPDB standard in both cases. *M.* Morozovella; *A.* Acarinina; *Gk.* Globigerinatheka; *S.* Subbotina; *H.* Hantkenina; *C.* Clavigerinella; *Gs.* Guembelitrionides; *I.* Igorina.

preservation of foraminiferal shells appears excellent on a micro-metre scale. Each clay sample was washed and sieved and mono-specific splits were prepared, typically of 5–10 adult individuals but more in the case of smaller species. Shells were ultrasonically cleaned, and treated with hydrogen peroxide to remove organic matter. Stable isotope analyses were conducted on a VG Prism mass spectrometer with a minimum analytical precision of $\pm 0.08\text{‰}$ for oxygen and $\pm 0.06\text{‰}$ for carbon. All data are available as Supplementary Information.

Multi-species results for each sample are given in Fig. 3. We have chosen not to apply a latitude-based ‘correction’ for seawater $\delta^{18}\text{O}$ by analogy with the modern ocean³⁴ because the precise form of the correction depends on unknown climatic factors^{17,35}, and because the outer continental shelf environments represented in this study were probably affected by boundary currents, making $\delta^{18}\text{O}$ values measured there unrepresentative of open-ocean values at the same latitudes. Because even un-encrusted planktonic foraminifera dwelling in the surface mixed-layer can precipitate some calcite in slightly deeper conditions, we take the most negative $\delta^{18}\text{O}$ value in the assemblage as representative of the seasonal maximum SST. Note that the benthic foraminifera in these samples reflect bottom-water temperatures on the continental shelf, and therefore are not comparable to coeval temperatures in deep-sea environments.

A possible objection to a stable isotope study in an outer-shelf environment is that it may be affected by freshwater inputs from large rivers, biasing the $\delta^{18}\text{O}$ measurements towards light (‘warm’) values. We do not consider this a significant problem for the Cretaceous and Middle Eocene samples analysed in this study for several reasons. The setting is a deep-water hemipelagic environment. Palaeogeographic reconstructions^{36,37} suggest a shoreline approximately 30–50 km inland of the present coastal localities

and, as today, an absence of large rivers. Trace fossils of the deep-water *Nereites* assemblage are common in parts of the succession³⁷ and planktonic microfossils are fully diverse, unlike modern environments affected by plumes of river water. Modern offshore Tanzania has normal marine salinities and temperatures due to the southward-flowing Somali current, which brings equatorial Indian Ocean water on to the narrow shelf, and similar circulation patterns and salinities have probably occurred since at least the Late Cretaceous¹⁵. The distribution of isotopic values is indicative of a well stratified open-ocean environment with no anomalous data as might be evident from a freshwater plume. Finally, freshwater runoff in tropical environments generally has an isotopic composition closer to that of sea water than in the middle and high latitudes, making it difficult to greatly affect the $\delta^{18}\text{O}$ without a very large freshwater influence. Despite this, we acknowledge that the Late Eocene samples from Tanzania and Alabama probably represent relatively shallower-water environments on account of the lower planktonic:benthic ratios and less diverse plankton assemblages, and so might be more likely to be affected by low salinities.

A single data set from a Late Cretaceous sample (sample LIN 99-100, *Abathomphalos mayaroensis* Zone, Maastrichtian age, 67 ± 2 Myr ago) is presented in Fig. 3a. Species have not been divided *a priori* into depth habitats because too little is known of Late Cretaceous planktonic foraminifer ecology. Nevertheless, our data are generally consistent with previous isotopic rankings^{20,21} in that small species belonging to the genera *Rugoglobigerina* and *Rugotruncana* and species with serially arranged chambers generally yield the more negative $\delta^{18}\text{O}$ values, whereas the larger, more ornate forms give more positive values. Two notable exceptions to this are *A. mayaroensis* which elsewhere²¹ has been recorded as a ‘deep dweller’ but has relatively negative $\delta^{18}\text{O}$ in our study, and *Rugoglobigerina rugosa* which has more positive isotopic ratios

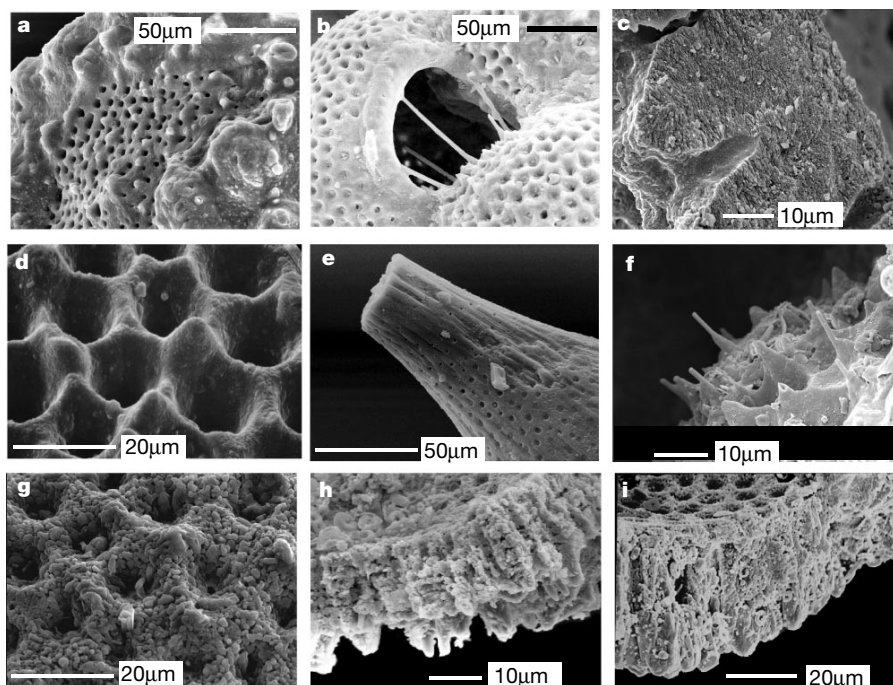


Figure 2 Scanning electron micrographs of planktonic foraminifera, contrasting well preserved (a–f) and poorly preserved (g–i) shell textures. **a**, *Globotruncana stuarti*, Maastrichtian, sample LIN 99-100, Lindi, Tanzania. **b**, *Globigerinatheka index* with preserved spines, Middle Eocene, sample LIN 99-17, Kitunda, Tanzania. **c**, *Globoquadrina tripartita*, Middle Eocene, cross-section of wall showing microgranular texture, sample LIN 99-17, Kitunda, Tanzania. **d**, *Globoquadrina tapuriensis*, holotype, lowermost Oligocene, detail of hexagonal pore-pits, sample FCRM-1645, Kitunda, Tanzania. **e**, *Hantkenina lehneri*, Middle Eocene, detail of tubulospine base, type section of Guayabal formation,

Vera Cruz, Mexico. **f**, *Guembeltrioides nuttalli*, Middle Eocene, spine bases, 1,170 m below sea floor, Hole IM-5, Adriatic Sea, Middle Eocene. **g**, *Globoquadrina tripartita*, Late Eocene, detail of recrystallized pore-pits, ODP sample 1052B, 10H-2, 133–135 cm in core section, western North Atlantic. **h**, *Acarinina praetopilensis*, Middle Eocene, cross-section of wall with coarse recrystallization, ODP Hole 865B, Pacific Ocean. **i**, *Morozovella aragonensis*, Middle Eocene, cross-section of recrystallized wall, DSDP Site 523, sample 49-2, 38–40 cm in core section.

than expected. We note that the relatively positive $\delta^{13}\text{C}$ of *Hedbergella holmdelensis* may indicate an association with symbiotic dinoflagellates. The most negative $\delta^{18}\text{O}$ in the assemblage is given by *Pseudotextularia elegans*, which has previously been recognized as a marker for surface-water conditions²¹.

The 'isotopic niches' of Middle Eocene species (Fig. 3b–f) are better known, and most species in the Tanzanian, Mexican and Adriatic samples plot in approximately their expected relative positions in comparison to earlier work^{26,38,39}. We note that the more encrusted mixed-layer species in the genera *Morozovella*, *Acarinina* and *Globigerinathea* tend to have more positive $\delta^{18}\text{O}$ values than do the more weakly ornamented acarininids, suggesting that a component of deeper water calcite added at gametogenesis is

present in those forms. We also note the unusual but consistent isotopic characteristics of the little-studied planispiral species *Pseudohastigerina micra*. We interpret this species as a mixed-layer calcifier that exhibits a large negative $\delta^{13}\text{C}$ vital effect, possibly because of incorporation of some metabolic CO_2 into the shell. More surprising was the relatively light $\delta^{18}\text{O}$ values of *Subbotina eocaena* in sample KIL 99-41, which suggests that not all Middle Eocene subbotinids were deep dwellers—as has sometimes been assumed¹⁸.

The Late Eocene assemblages (Fig. 3g–i) show smaller isotopic differentials between species, especially in $\delta^{13}\text{C}$, as has frequently been observed in other studies³⁸. Surface-water conditions seem to be indicated by the genera *Pseudohastigerina*, *Globigerinathea* and

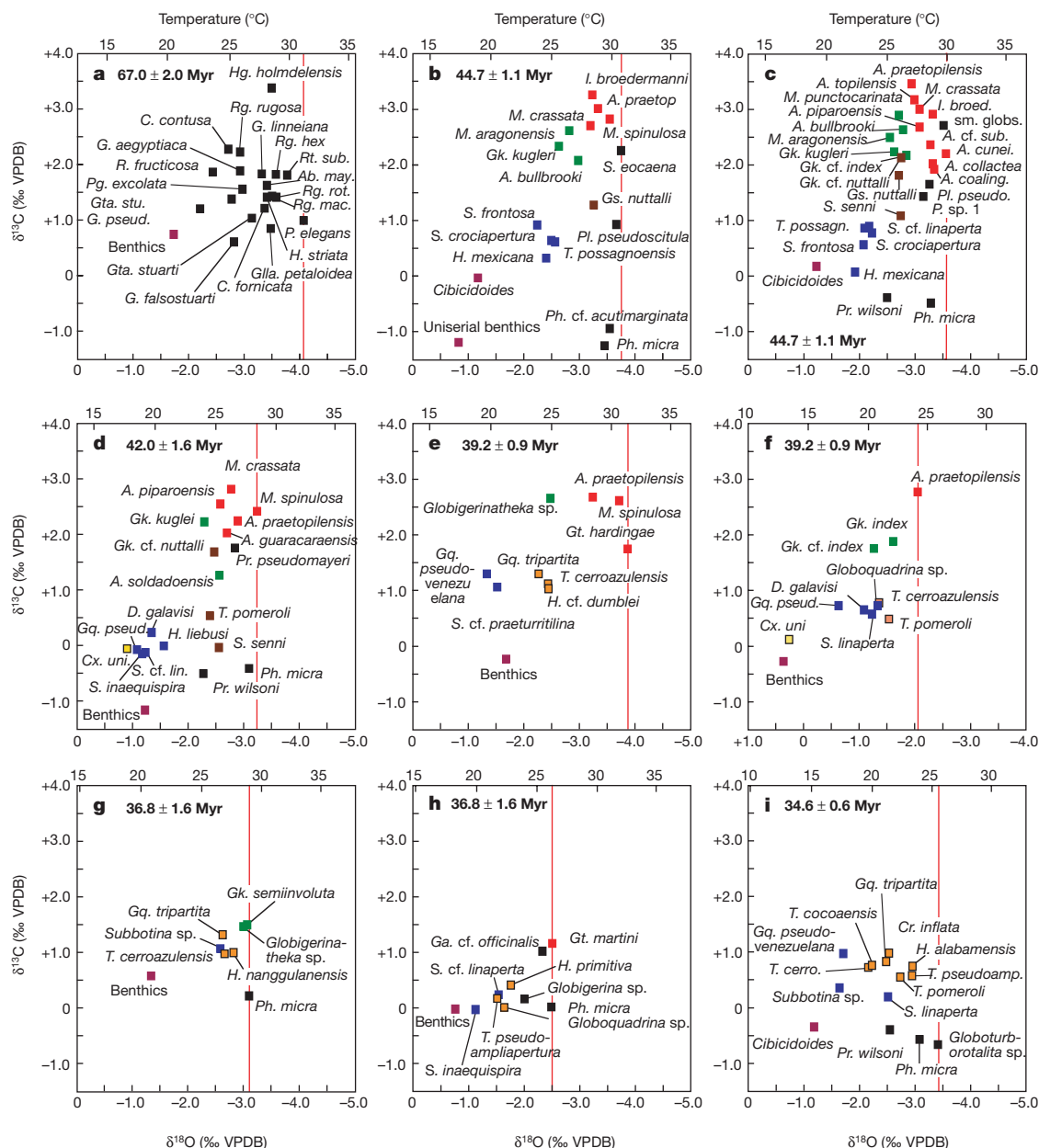


Figure 3 Multi-species stable isotope arrays for a variety of exceptionally well preserved samples. Vertical red lines indicate the most negative $\delta^{18}\text{O}$ of each assemblage, indicating inferred sea surface temperature. The temperature scale is calculated using the equation of ref. 50, and includes an ice-volume correction to seawater $\delta^{18}\text{O}$ of -1.00‰ for the Late Cretaceous, -0.75‰ for the Middle Eocene, and -0.5‰ for the Late Eocene. Samples are dated to within one planktonic foraminifer biozone. Coloured squares indicate habitat as follows: red, mixed-layer symbiotic; green, encrusted mixed-layer

symbiotic; brown, intermediate habitat; blue, thermocline; yellow, sub-thermocline; black, planktonic, habitat uncertain; purple, benthic. See Supplementary Information for biostratigraphy and full Linnaean binomials of foraminifer species. **a**, Sample LIN 99-100, Lindi, Tanzania. **b**, Sample KIL 99-41, Kilwa Masoko, Tanzania. **c**, Sample RAS 99-17, Ras Mtama, Tanzania. **d**, Guayabal formation, Mexico. **e**, Sample LIN 99-17, Lindi, Tanzania. **f**, Istra-more 5 well, Adriatic Sea. **g**, Sample PP98-L11, Lindi, Tanzania. **h**, Cocoa Sands formation, Alabama. **i**, Sample NAM 99-01, Namadingura, Tanzania.

the small globular genus *Globoturborotalita* which was probably ancestral to many modern mixed-layer species.

Diagenesis reconsidered

Although the relative positions of the various species in our sample are generally as expected from earlier work, a greater range of values is observed in both stable isotopes, and the $\delta^{18}\text{O}$ data are considerably more negative than has generally been reported. These discrepancies can be resolved if we interpret our exceptionally well preserved material as representing the true, warm, SSTs, whereas other more recrystallized material has been 'overprinted' by a component of diagenetic calcite precipitated in cold bottom waters. This accords with the fact that $\delta^{18}\text{O}$ values are expected to be altered toward more positive values in early marine diagenesis^{3,5,6}, which is the reverse of the more normal trend towards negative values in meteoric and high-temperature diagenesis.

To explore this possibility, particularly with regard to the problem that carbon isotope differentials between species are apparently retained to some extent during diagenesis³, we have plotted (Fig. 4) the data from a representative sample from our study (RAS 99-17) on the same scale as the DSDP Site 523 data from Fig. 1 (which is approximately the same age). A total of 11 species are common to both data sets, and are joined by lines for clarity. The burial depth of the Site 523 samples is approximately 200 m, within the zone at which diagenesis is expected to cause a shift towards positive $\delta^{18}\text{O}$ (ref. 5). We have plotted on Fig. 4 the expected isotopic composition of diagenetic calcite at shallow burial depths³. We note a clear convergence of the lines towards this composition, which at face value would indicate a likely component of 50% diagenetic calcite and 50% original calcite in the Site 523 data, and a reduction of apparent palaeotemperature of nearly 15 °C.

However, we acknowledge that the RAS 99-17 temperatures may not be representative of the Site 523 assemblage before diagenesis. The Tanzanian data set (palaeolatitude ~19° S) probably existed in a warm boundary current, whereas the Site 523 data are from a higher sub-tropical palaeolatitude (~32° S) and may have been influenced by a colder northward proto-Benguela current. If the RAS 99-17 data were shifted by several degrees to cooler values to compensate, the data would still yield a plausible set of 'mixing lines', still indicating a substantial overprint (~50% by volume) in both $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$, but converging on warmer bottom-water temperatures. That interpretation would in fact be more in line with benthic foraminifer stable isotope data from the Middle

Eocene that indicate bottom-water temperatures nearer 10 °C (ref. 40).

We infer from this that most planktonic foraminifer stable isotope data from carbonate oozes and chalks are 'suspect', and may represent a roughly equal combination of surface- and bottom-water signals. The effect probably varies in degree between sites according to their burial depths and diagenetic histories. Samples from more clay-rich sites and horizons are probably less altered than more carbonate-rich sites. This result would have profound importance for palaeoceanography in general, including studies of the long-term evolution of global climate and the interpretation of transient climate events which may involve changes in both surface- and bottom-water conditions.

The 'broad warm tropics' orthodoxy?

In Fig. 5 we show oxygen isotope data from mixed-layer and deeper-dwelling foraminifera from a variety of locations plotted as a function of latitude, with the new data from excellently preserved carbonates highlighted in red. The contrast between the new and old data is evident, in both the absolute values and the surface-to-thermocline differentials.

A corollary of our interpretations is that diagenesis would be expected not only to decrease the measured temperatures and inter-species differentials, but also to reduce the apparent meridional temperature gradients inferred from recrystallized material. This is because deep-sea temperatures and early diagenetic conditions are expected to be fairly homogeneous at any given time, regardless of latitude. We therefore propose that meridional gradients were probably steeper than previously thought in the Late Cretaceous and Eocene epochs. This is supported by a comparison of the tropical data with the higher-latitude sample from the Adriatic, which indicates markedly cooler temperatures for the same species. However, we highlight the need for more data from excellently preserved material from a variety of latitudes to test this proposition.

Present-day SSTs rarely exceed 30 °C (except in isolated basins), leading some to suggest that negative feedbacks involving cloud cover⁴¹, evaporation or changes in circulation patterns⁴² limit open ocean temperatures. Several of our estimates are above 30 °C, and it is unlikely that we have sampled the warmest parts of the oceans. So although we acknowledge multiple uncertainties in estimating temperatures for distant time periods, our data would appear to contradict the existence of a strict homeostatic limit to tropical SSTs in greenhouse conditions.

Low-latitude temperatures have a significant effect on wind strengths, monsoons, oceanic upwelling intensities and zonal patterns of precipitation and evaporation. A recent comparison³⁵ of Eocene climate simulations with cool and warm (>30 °C) tropical SSTs suggests that a broad warm tropical belt produces a realistic simulation of Eocene climate that is consistent with most geological evidence³⁵. It has also been pointed out¹⁷ that even moderately higher tropical temperatures than today (+2 °C) would markedly increase evaporation rates and hence latent heat transport by the atmosphere, helping to explain the inferred polar warmth of past warm climate phases.

The warm tropics view is also supported by the geographical distributions of temperature-sensitive organisms such as larger foraminifera, corals, mangroves, palms and reptiles. All of these organisms have present-day minimum temperature requirements that have previously been argued as contradicting the cool tropics view⁴³⁻⁴⁵. Unfortunately, it is more difficult to place a precise upper bound on the likely temperatures given the distributions of fossil organisms, and the question has received less attention⁴⁶. Symbiotic corals may be the most sensitive of these groups to high temperatures, and it is interesting that they were apparently less common and diverse in Cretaceous and Palaeogene times than they are today⁴⁶. It has been suggested that Pacific carbonate platforms

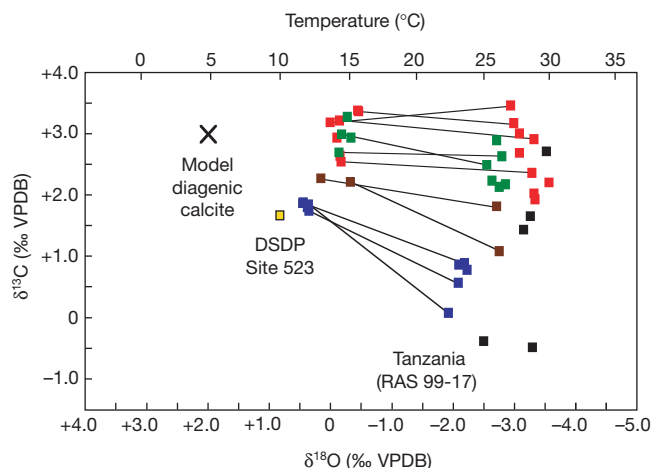


Figure 4 Comparison of stable isotope data from a typical recrystallized sample (DSDP Site 523, data as in Fig. 1) with a representative sample from this study of approximately the same age (sample RAS 99-17). Lines join species that are common to the two data sets. The approximate composition of diagenetic calcite is from ref. 3. Colours as in Fig. 3.

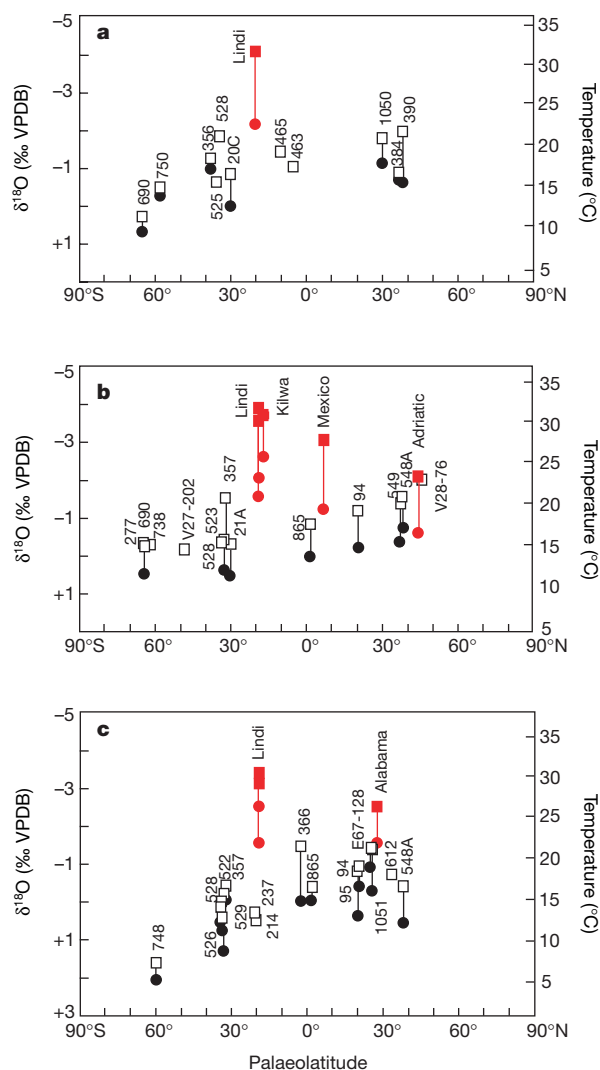


Figure 5 Zonal profiles of planktonic foraminifer $\delta^{18}\text{O}$ as a function of palaeolatitude. **a**, *A. mayaroensis* zone, Late Cretaceous epoch; **b**, zone P11–P14, Middle Eocene; **c**, zone P15–P16, Late Eocene. Squares, mixed-layer-dwelling species; circles, deeper-dwelling species. Data from this study are highlighted in red. Ice-volume corrections for seawater $\delta^{18}\text{O}$ are the same as in Fig. 2.

that moved on the plate tectonic conveyor into the tropics during these times often died, possibly as a result of high temperatures⁴⁷. Also suggestive is that organisms such as crocodilians, which peak in diversity in the modern equatorial region, had sub-tropical diversity peaks in both hemispheres during greenhouse periods⁴⁵, possibly because they were excluded from the equator by unfavourable climate conditions.

Discussion

Our analysis of planktonic foraminifer shells from hemipelagic clays suggests that, for those parts of the Late Cretaceous and Eocene epochs that we have sampled, tropical temperatures were at least as warm as today, and probably several degrees warmer. This result is in line with known patterns of biotic distribution and the predictions of climate models, and allows us to dispose of the 'cool tropic paradox' that has bedevilled the study of past warm climates. The new evidence also removes a significant impediment²² to the view that high levels of greenhouse gases contributed to global warming during these periods^{48,49}.

We contend that most previous workers, including ourselves, have been misled to some extent by fine-scale recrystallization of

planktonic foraminifer shells, which occurs at shallow burial depths in open ocean pelagic oozes and chalks. This process introduces a much larger component of diagenetic calcite than has generally been recognized, making such shells unsuitable for sea surface palaeo-temperature analysis. However, we acknowledge that even if a recrystallized shell is a roughly equal mixture of sea-surface and early diagenetic calcite, much work can still be done, within a modified interpretative framework, in the study of long-term climate change and also in the study of Milankovitch-scale variation and transient events. Also, we stress that benthic foraminifer shells are probably not affected by diagenesis to the same degree, so the classic deep-sea $\delta^{18}\text{O}$ record of the Cenozoic era⁴⁰ remains a valuable benchmark for long-term palaeoclimate research. As scientific ocean drilling enters a new phase, we hope that our data will act as a spur for the targeting of hemipelagic muds of Cretaceous and Palaeogene age from a variety of latitudes that are likely to contain exceptionally well preserved planktonic foraminiferal shells. □

Received 29 May; accepted 14 August 2001.

1. Urey, H. C. The thermodynamic properties of isotopic substances. *J. Chem. Soc.* **1947**, 562–581 (1947).
2. Emiliani, C. Depth habitats of some species of pelagic foraminifera as indicated by oxygen isotope ratios. *Am. J. Sci.* **252**, 149–158 (1954).
3. Crowley, T. J. & Zachos, J. C. in *Warm Climates in Earth History* (eds Huber, B. T., MacLeod, K. S. & Wing, S. C.) 50–76 (Cambridge Univ. Press, Cambridge, 2000).
4. Spero, H. J., Bijma, J., Lea, J. W. & Bemis, B. E. Effect of seawater carbonate chemistry on planktonic foraminiferal carbon and oxygen isotope values. *Nature* **390**, 497–500 (1997).
5. Schrag, D. P., DePaolo, D. J. & Richter, F. M. Reconstructing past sea surface temperatures: Correcting for diagenesis of bulk marine carbonate. *Geochim. Cosmochim. Acta* **59**, 2265–2278 (1995).
6. Schrag, D. P. Effects of diagenesis on the isotopic record of late Paleogene tropical sea surface temperatures. *Chem. Geol.* **161**, 215–244 (1999).
7. Shackleton, N. J. & Boersma, A. The climate of the Eocene ocean. *J. Geol. Soc. Lond.* **138**, 153–157 (1981).
8. Savin, S. M. The history of earth's surface temperature during the past 100 million years. *Annu. Rev. Earth Planet. Sci.* **5**, 319–355 (1977).
9. Shackleton, N. J. in *Fossils and Climate* (ed. Brenchley, P.) 27–34 (Wiley, Chichester, 1984).
10. Barron, E. J. Eocene equator-to-pole surface ocean temperatures: a significant climate problem? *Paleoceanography* **2**, 729–739 (1987).
11. Rind, D. & Chandler, M. Increased ocean heat transports and warmer climate. *J. Geophys. Res.* **96**, 7437–7461 (1991).
12. Horrell, M. A. Energy balance constraints on ^{18}O based paleo-sea surface temperature estimates. *Paleoceanography* **5**, 339–348 (1990).
13. Manabe, S. & Bryan, K. CO_2 -induced change in a coupled ocean-atmosphere model and its paleoclimatic implications. *J. Geophys. Res.* **90**, 11689–11708 (1985).
14. Sloan, L. C. & Rea, D. K. Atmospheric carbon dioxide and early Eocene climate: A general circulation model sensitivity study. *Paleoceanogr., Palaeoclimatol. Palaeoecol.* **119**, 275–292 (1995).
15. Bush, A. B. G. & Philander, S. G. H. The late Cretaceous: Simulation with a coupled atmosphere-ocean general circulation model. *Paleoceanography* **12**, 495–516 (1997).
16. Rind, D. Latitudinal temperature gradients and climate change. *J. Geophys. Res.* **103**, 5943–5971 (1998).
17. Bice, K. L., Scotese, C. R., Seidov, D. & Barron, E. J. Quantifying the role of geographic change in Cenozoic ocean heat transport using uncoupled atmosphere and ocean models. *Paleoceanogr. Palaeoclimatol. Palaeoecol.* **161**, 295–310 (2000).
18. Bralower, T. et al. Late Paleocene to Eocene paleoceanography of the equatorial Pacific Ocean: Stable isotopes recorded at Ocean Drilling Program Site 865, Allison Guyot. *Paleoceanography* **10**, 841–865 (1995).
19. Wade, B. in *Western North Atlantic Paleogene and Cretaceous Paleocceanography* (eds Kroon, D., Norris, R. D. & Klaus, A.) 273–291 (Spec. Publ. 183, Geological Society, London, 2001).
20. D'Hondt, S. & Arthur, M. A. Late Cretaceous oceans and the cool tropic paradox. *Science* **271**, 1838–1841 (1996).
21. D'Hondt, S. & Arthur, M. A. Interspecies variation in stable isotopic signals of Maastrichtian planktonic foraminifera. *Paleoceanography* **10**, 123–125 (1995).
22. Veizer, J., Goddard, Y. & François, L. M. Evidence for decoupling of atmospheric CO_2 and global climate during the Phanerozoic eon. *Nature* **408**, 698–701 (2000).
23. Crowley, T. J. & Berner, R. A. CO_2 and climate change. *Science* **292**, 870–872 (2001).
24. Killingley, J. S. Effects of diagenetic recrystallization on $^{18}\text{O}/^{16}\text{O}$ values of deep-sea sediments. *Nature* **301**, 594–597 (1983).
25. Wilson, P. A. & Opdyke, B. N. Equatorial sea surface temperatures for the Maastrichtian revealed through remarkable preservation of metastable carbonate. *Geology*, **24**, 553–558 (1996).
26. Pearson, P. N., Shackleton, N. J. & Hall, M. A. Stable isotope paleoecology of middle Eocene planktonic foraminifera and multi-species isotope stratigraphy, DSDP Site 523, South Atlantic. *J. Foram. Res.* **23**, 123–140 (1993).
27. Andreasson, F. P. & Schmitz, B. Tropical Atlantic seasonal dynamics in the early middle Eocene from stable oxygen and carbon isotope profiles of mollusk shells. *Paleoceanography*, **13**, 183–192 (1998).
28. Kolodny, Y. & Raab, M. Oxygen isotopes in phosphatic fish remains from Israel: paleothermometry of tropical Cretaceous and Tertiary shelf waters. *Paleoceanogr. Palaeoclimatol. Palaeoecol.* **64**, 59–67 (1988).
29. Norris, R. D. & Wilson, P. A. Low-latitude sea-surface temperatures for the mid-Cretaceous and the evolution of planktic foraminifera. *Geology* **26**, 823–826 (1998).

30. Wilson, P. A. & Norris, R. D. Warm tropical ocean surface and global anoxia during the mid-Cretaceous period. *Nature* **412**, 425–429 (1991).
31. Hemleben, C., Spindler, M. & Anderson, O. R. *Modern Planktonic Foraminifera* (Springer, New York, 1989).
32. Blow, W. H. & Banner, F. T. in *Fundamentals of Mid-Tertiary Stratigraphical Correlation* (ed. Eames, F. T.), 61–151 (Cambridge Univ. Press, Cambridge, 1962).
33. Blow, W. H. *The Cainozoic Globigerinida: a Study of the Morphology, Taxonomy and Evolutionary Relationships and the Stratigraphical Distribution of some Globigerinida (mainly Globigerinacea)* Vols 1–3 (Brill, Leiden, 1979).
34. Zachos, J. C., Stott, L. D. & Lohmann, K. C. Evolution of early Cenozoic marine temperatures. *Paleoceanography* **9**, 353–387 (1994).
35. Huber, M. & Sloan, L. C. Climatic response to tropical sea surface temperature changes on a 'greenhouse' Earth. *Paleoceanography* **15**, 443–450 (2000).
36. Kent, P. E., Hunt, M. A. & Johnstone, M. A. *The Geology and Geophysics of Coastal Tanzania*. (Natural Environment Research Council Geophysical Paper no. 6, HMSO, London, 1971).
37. Schluter, T. *Geology of East Africa* 239 (Borntraeger, Stuttgart, 1997).
38. Boersma, A., Premoli Silva, I. & Shackleton, N. J. Atlantic Eocene planktonic foraminiferal paleohydrographic indicators and stable isotope paleoceanography. *Paleoceanography* **2**, 287–331 (1987).
39. Coxall, H. K., Pearson, P. N., Shackleton, N. J. & Hall, M. A. Hantkeninid depth adaptation: an evolving life strategy in a changing ocean. *Geology* **28**, 87–90 (2000).
40. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65Ma to present. *Science* **292**, 686–693 (2001).
41. Ramanathan, V. & Collins, W. Thermodynamic regulation of ocean warming by cirrus clouds deduced from satellite observations of the 1987 El Niño. *Nature* **351**, 27–32 (1991).
42. Waliser, D. E. Some considerations on the thermostat hypothesis. *Bull Am. Meteorol. Soc.* **77**, 357–360 (1996).
43. Adams, C. G., Lee, D. E. & Rosen, B. R. Conflicting isotopic and biotic evidence for tropical sea-surface temperatures during the Tertiary. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **77**, 289–313 (1990).
44. Graham, A. Neotropical Eocene coastal floras and $^{18}\text{O}/^{16}\text{O}$ -estimated warmer vs. cooler equatorial waters. *Am. J. Bot.* **81**, 301–306 (1994).
45. Markwick, P. Fossil crocodilians as indicators of Late Cretaceous and Cenozoic climates: implications for using palaeontological data in reconstructing palaeoclimate. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **137**, 207–271 (1998).
46. Barron, E. J. in *Effects of Past Global Change on Life* 108–117 (National Academy Press, Washington D.C., 1995).
47. Wilson, P. A., Jenkyns, H. C., Elderfield, H. & Larson, R. L. The paradox of drowned carbonate platforms and the origin of Cretaceous Pacific guyots. *Nature* **392**, 889–894 (1998).
48. Pearson, P. N. & Palmer, M. R. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature* **406**, 695–699 (2000).
49. Retallack, G. J. A 300-million-year record of atmospheric carbon dioxide from fossil plant cuticles. *Nature* **411**, 287–290 (2001).
50. Erez, B. & Luz, J. Experimental paleotemperature equation for planktonic foraminifera. *Geochim. Cosmochim. Acta* **47**, 1025–1031 (1983).

Supplementary information is available on *Nature's* World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank the Tanzania Petroleum Development Corporation and A. Karega for field support, and P. Sexton and J. Whittaker for providing SEM images. We also thank V. Premec-Fucek and R. Norris for providing samples for analysis, and C. Hemleben for insights regarding processes of biomineralization and diagenesis in planktonic foraminifer shells.

Correspondence and requests for materials should be addressed to P.N.P. (e-mail: paul.pearson@bristol.ac.uk).

Mutations in a member of the *ADAMTS* gene family cause thrombotic thrombocytopenic purpura

Gallia G. Levy*, William C. Nichols†, Eric C. Lian‡, Tatiana Foroud§, Jeanette N. McClintick§, Beth M. McGee*, Angela Y. Yang*, David R. Siemieniak*, Kenneth R. Stark*, Ralph Grupp||, Ravindra Sarode¶, Susan B. Shurin#, Visalam Chandrasekaran☆, Sally P. Stabler**, Hernan Sabio††, Eric E. Bouhassira‡‡, Jefferson D. Upshaw, Jr.§§, David Ginsburg* & Han-Mou Tsai|||

* Howard Hughes Medical Institute, Departments of Internal Medicine and Human Genetics, and Cellular and Molecular Biology Program, University of Michigan Medical Center, Ann Arbor, Michigan 48109, USA

† Division of Human Genetics, Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA

‡ Hemophilia and Thrombosis Center and Sylvester Cancer Center, University of Miami and Veterans Affairs Medical Center, Miami, Florida 33136, USA

§ Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA

|| Division of Hematology/Oncology, Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA

¶ Blood Bank, University of Texas Southwestern School of Medicine, Dallas, Texas 75390, USA

Department of Pediatrics, Rainbow Babies and Children's Hospital, and the Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, USA

☆ Blood Bank, Long Island Jewish Medical Center, and Albert Einstein College of Medicine, New Hyde Park, New York 11040, USA

** Department of Medicine, University of Colorado Health Sciences Center, Denver, Colorado 80262, USA

†† Pediatric Hematology/Oncology, Medical College of Georgia, Augusta, Georgia 30912, USA

‡‡ Department of Medicine/Hematology and Department of Cell Biology, Albert Einstein College of Medicine, Bronx, New York 10461, USA

§§ The Memphis Cancer Center, Memphis, Tennessee 38119, USA

||| Division of Hematology, Montefiore Medical Center, and Albert Einstein College of Medicine, Bronx, New York 10467, USA

Thrombotic thrombocytopenic purpura (TTP) is a life-threatening systemic illness of abrupt onset and unknown cause. Proteolysis of the blood-clotting protein von Willebrand factor (VWF) observed in normal plasma is decreased in TTP patients. However, the identity of the responsible protease and its role in the pathophysiology of TTP remain unknown. We performed genome-wide linkage analysis in four pedigrees of humans with congenital TTP and mapped the responsible genetic locus to chromosome 9q34. A predicted gene in the identified interval corresponds to a segment of a much larger transcript, identifying a new member of the *ADAMTS* family of zinc metalloproteinase genes (*ADAMTS13*). Analysis of patients' genomic DNA identified 12 mutations in the *ADAMTS13* gene, accounting for 14 of the 15 disease alleles studied. We show that deficiency of *ADAMTS13* is the molecular mechanism responsible for TTP, and suggest that physiologic proteolysis of VWF and/or other *ADAMTS13* substrates is required for normal vascular homeostasis.

TTP is characterized by intravascular destruction of erythrocytes and blood platelets resulting in anaemia and thrombocytopenia. Diffuse platelet-rich microthrombi are observed in small blood vessels of multiple organs, with the major complications including renal failure and neurologic dysfunction. In the related disorder, haemolytic uraemic syndrome (HUS), neurologic symptoms are less evident and renal failure is more prominent. HUS is most frequently observed in young children, usually following infection with a specific Shiga-toxin-producing strain of *Escherichia coli*. Although TTP-like disorders have also been associated with various medications, bone marrow transplantation, pregnancy, human immunodeficiency virus (HIV) infection, and autoimmune disease, most cases appear sporadically, without an obvious precipitating event. A subset of patients from these sporadic cases experience a chronic relapsing course. Rare familial cases of TTP have also been reported, generally associated with neonatal onset and frequent relapses (OMIM (<http://www.ncbi.nlm.nih.gov/Omim>) (accession numbers 274150 and 276850)^{1,2}.

Before the development of modern treatment protocols consisting of fresh-frozen plasma infusion with or without plasma exchange, fatality during an acute episode of TTP was >90% (refs 2 and 3). Although treatment outcome has improved significantly, the molecular pathogenesis of TTP is still unknown and the specific plasma factor(s) responsible for the acute onset of this disease, or recovery after treatment, remains to be identified. Unusually large multimeric forms of the platelet-adhesive blood-coagulation

protein VWF have been observed in the plasma of TTP patients and are proposed to have a pathogenic role in the formation of the microvascular VWF- and platelet-rich thrombi characteristic of this disorder⁴. Consistent with this hypothesis, a proteolytic activity that degrades large VWF multimers to smaller sizes^{5,6} has been shown to be decreased in the plasma of TTP patients, either on the basis of a constitutional deficiency in congenital cases or the presence of an autoantibody inhibitor in most sporadic adult-onset cases^{7–10}. These findings suggest that TTP may be triggered by accumulation of large, highly adhesive VWF multimers in the absence of physiologic processing by this VWF-cleaving protease. However, other studies have implicated platelet-aggregating proteins^{11,12} or endothelial injury¹³ as the underlying mechanism, and enhanced rather than decreased VWF proteolysis has been reported in some patients¹⁴. Although the VWF-cleaving protease has been partially purified^{5,6}, it seems to be present at relatively low levels in plasma and its identification at the sequence level has remained elusive.

VWF-cleaving protease activity is a semidominant trait

To further explore the molecular pathogenesis of TTP, we used a genetic approach to study the four pedigrees shown in Fig. 1. Activity of VWF-cleaving protease was measured in the plasma of the seven affected individuals, and was 2–7% of normal (0.02–0.07 U ml⁻¹); none of the patients tested positive for inhibitors of the protease. Plasma levels of the protease in the parents of the affected individuals were 0.51–0.68 U ml⁻¹, consistent with a heterozygous

carrier state. Similarly, levels for at-risk siblings of the patients and parents fell into a bimodal distribution, with one peak consistent with carriers and the other indistinguishable from the normal distribution (Fig. 2). These results demonstrate that the protease activity assay used here reliably distinguishes between normal and carrier individuals in these families. This observation suggested that the plasma level of VWF-cleaving protease could be used as a phenotypic trait for linkage analysis to map the corresponding locus, providing considerably greater genetic power than would be available from analysis of the clinical phenotype alone.

Gene for familial TTP maps to chromosome 9q34

A genome-wide linkage scan was performed with 382 polymorphic microsatellite markers to analyse DNA from affected individuals and other informative family members (Fig. 1). Two-point linkage analysis with a recessive model gave a maximum lod (log likelihood ratio) score of 2.36 at $\theta = 0.0$ for marker D9S164 on chromosome 9q34. A lod score of 5.63 at $\theta = 0.0$ was obtained for the same marker with a codominant model. There was still substantial evidence of linkage to D9S164 when a codominant model with 95% penetrance was used (lod score of 4.12 with $\theta = 0.0$ at D9S164). Multipoint analysis for D9S164 and four flanking markers yielded a maximum lod score of 7.37 at marker D9S164. Genotypes for seven additional markers in this region^{15,16} allowed the gene to be placed in

the ~7-centimorgan (cM) interval between markers D9S1863 and D9S1818 (Figs 1 and 3a). Analysis of new polymorphic markers (see Supplementary Information Table 3) designed using simple sequence-repeat data available from the Human Genome Project working draft (<http://genome.ucsc.edu>)¹⁷ further narrowed the candidate interval to an ~2.3-megabase (Mb) genomic segment between markers GL2-1 and D9S1818. In all but one case, carrier status as determined by haplotype analysis was consistent with the phenotypic designation according to plasma protease level. The exception, individual II2 in pedigree A, shares the affected haplotype of her brother (II4), but has a protease level of 0.8 U ml⁻¹, which is borderline between the normal and carrier ranges.

Gene for familial TTP encodes an ADAMTS metalloproteinase

Analysis of the candidate interval using public genome databases (<http://genome.ucsc.edu>, <http://www.ncbi.nlm.nih.gov>)¹⁷ identified ~20 known or predicted genes (Fig. 3a). Initial attention focused on genes likely to encode a protease or protease cofactor. *Ficolin 2* (*FCN2*) mapped to distal chromosome 9 but could not be identified in available BAC sequence from the candidate interval. However, in light of previous reports suggesting a protease-associated function for some ficolin family members¹⁸, and the possibility that *FCN2* might lie in one of the three large genomic sequence gaps shown in Fig. 3a, the coding exons and intron-exon

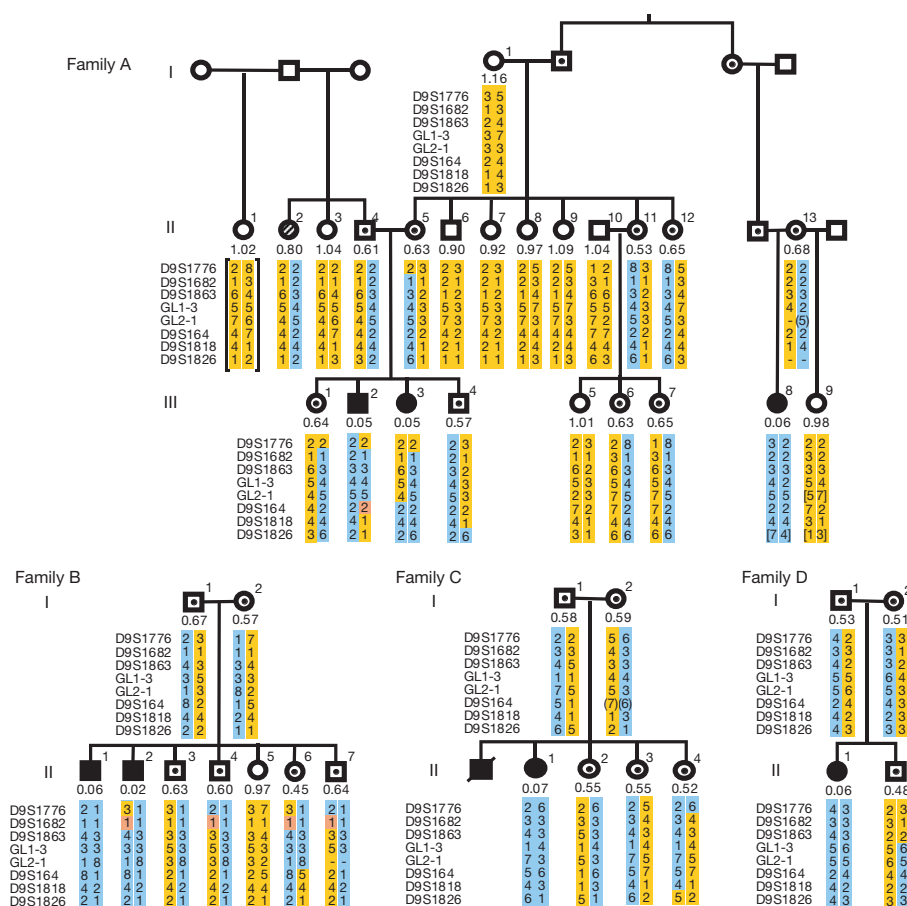


Figure 1 Pedigrees used for linkage analysis. VWF-cleaving protease levels (in units per millilitre) are shown beneath the symbol for each individual. Circles indicate females and squares indicate males. Affected individuals are indicated by solid symbols and carriers by symbols with dots. Individual AII-2 was classified as phenotype unknown (hatched symbol) on the basis of an intermediate protease level (0.8). Seventeen markers (13 from the Genethon¹⁵ and Marshfield¹⁶ comprehensive genetic maps and four designed from sequence-repeat information available at the Human Genome Project working draft¹⁷ (see Supplementary Information Table 3)) were used for haplotype

analysis. Only selected markers are shown. Chromosomes carrying affected alleles are blue and normal chromosomes are yellow. Areas where recombination cannot be definitively assigned are orange. Only recombination events between affected and unaffected chromosomes are shown. Inferred genotypes are indicated in parentheses, and genotypes of unknown phase in square brackets. Recombination events in individuals AIII-3 and BII-6 place the responsible gene below marker GL2-1 and a recombination event in individual AIII-2 places the gene above marker D9S1818.

boundaries of this gene were amplified by polymerase chain reaction (PCR) from patients' DNA and subjected to sequence analysis. No candidate mutations were identified. Two putative genes in the candidate interval, *KIAA0605*, an uncharacterized expressed sequence tag (EST) from a brain complementary DNA library¹⁹, and the predicted open reading frame *C9ORF8*, exhibited homology to the ADAMTS family of metalloproteinases, but seemed to lack the conserved protease catalytic domain. Partial DNA sequence analysis of exons and flanking intron sequences failed to identify any mutations in *KIAA0605*. However, the identification of several candidate missense mutations in the predicted exons of *C9ORF8* led to further, more detailed analysis of this candidate gene. Through a combination of cDNA cloning, PCR with reverse transcription (RT-PCR), rapid amplification of cDNA ends (RACE) and genomic sequence analysis, the full-length *ADAMTS13* cDNA sequence (GenBank accession number AF414401) and corresponding genomic structure were deduced (Fig. 3b) and found to encode a complete, potentially catalytically active ADAMTS protease.

Members of the ADAM (a disintegrin and metalloproteinase) family are membrane-anchored proteases with diverse functions. Known members include fertilins α and β , implicated in sperm-egg fusion, and the 'shedases', such as TACE (TNF- α convertase), which mediate the shedding of cell-surface proteins²⁰. Members of the ADAMTS family are distinguished from ADAMs by the presence of one or more thrombospondin-1-like (TSP1) domain(s) at the carboxy terminus, which are thought to mediate interactions with components of the extracellular matrix^{21–24}. In addition, ADAMTS proteins lack the EGF repeat, transmembrane domain and cytoplasmic tail typically observed in ADAM metalloproteinases. Mutations in *ADAMTS2* (procollagen N-proteinase), the only family

member previously associated with a human genetic disease, result in the connective tissue disorder, Ehlers-Danlos syndrome type VII (ref. 25). Although *ADAMTS1* mutations have not been identified in humans, genetically deficient mice exhibit growth retardation, adipose tissue abnormalities, and fibrotic changes throughout the genitourinary system, suggesting a critical importance for ADAMTS1 in organogenesis and tissue remodelling²⁶. ADAMTS4 and ADAMTS5/11 (aggrecanases) cleave the core protein of the cartilage proteoglycan, aggrecan, and may be important in inflammatory joint disease²⁷. Procollagen II has recently been identified as the substrate for ADAMTS3 (ref. 28). The function and protein substrates for the remaining ADAMTS family members are unknown.

The *ADAMTS13* gene spans 29 exons encompassing ~37 kb in the human genome and encodes a protein with 1,427 amino acids (Fig. 3b). Analysis of RT-PCR and cloned cDNA sequences provided evidence for alternative splicing of exon 17 (GenBank accession number AF414400), resulting in a frameshift that predicts a truncated 842-amino-acid form of the protein lacking the six C-terminal TSP1 repeats. Comparative analysis with draft mouse genomic sequences demonstrates a high degree of conservation throughout the coding exons and identifies an additional potential exon located between the current exons 22 and 23, which may indicate another splice isoform. These findings suggest the potential for differentially regulated alternative isoforms of ADAMTS13 with diverse biological functions in addition to the putative proteolytic processing of VWF. Alternative splicing has been observed for two other ADAMTS genes, *ADAMTS2* (ref. 25) and *ADAMTS9* (ref. 23), resulting in similar variation in the number of C-terminal TSP1 repeats.

The domain structure of ADAMTS13 is depicted at the bottom of Fig. 3b. A predicted signal peptide is followed by a short propeptide domain ending in a potential propeptide convertase cleavage site at amino acids 71–74 (RQRR), suggesting that proteolytic processing, either in the *trans* Golgi or at the cell surface, is required for activation. The protease domain that follows contains a perfect match for the HEXGHXXGXXHD (where X is any amino acid) consensus sequence of the extended catalytic site shared between snake venom metalloproteinases and ADAM family members^{20,21,29}. The catalytic domain is followed by TSP1 and spacer domains characteristic of the ADAMTS family. An RGD sequence, present in only one other mature ADAMTS protein (*ADAMTS2*), is located immediately C terminal to the first TSP1 domain of ADAMTS13, suggesting a possible integrin interaction. The C terminus contains six additional TSP1 repeats, followed by a segment with homology to the CUB domain identified in a number of developmentally regulated proteins³⁰. The previously reported inhibitor profile and metal cation dependence of the VWF-cleaving protease^{5,6,31} are consistent with its identity as an ADAMTS. The predicted, non-glycosylated relative molecular mass of ADAMTS13 is 154,000 (M_r 154K), consistent with a previously estimated M_r of 200K for partially purified VWF-cleaving protease⁵, although considerably smaller than the 300K M_r reported by others⁶.

Northern blot analysis detected an *ADAMTS13* messenger RNA of ~4.7-kb specifically in the liver, with a truncated (~2.3 kb) mRNA faintly visible in the placenta (Fig. 4a). These data suggest that plasma VWF-cleaving protease may be derived primarily from *ADAMTS13* expression in the liver. The strong RT-PCR signal seen in the ovary, and variable expression in other tissues (Fig. 4b), may point to other potential functions for this protein. The absence of detectable transcripts in other highly vascular tissues such as the lung and heart may indicate that the vascular endothelium is not a primary site of *ADAMTS13* expression.

Mutations in *ADAMTS13* cause familial TTP

Analysis of DNA sequences identified mutations within the *ADAMTS13* gene in all four of the pedigrees depicted in Fig. 1, as well as in three additional TTP patients not included in the original

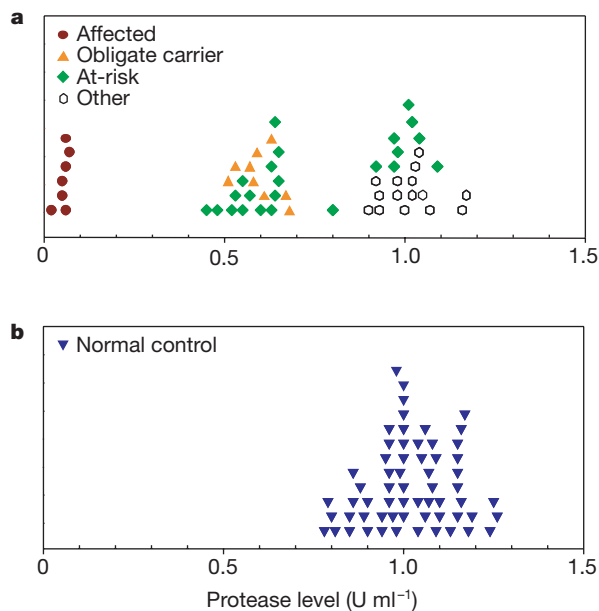


Figure 2 Plasma VWF-cleaving protease levels. **a**, Levels for all individuals shown in Fig. 1, as well as additional members of family A. **b**, Levels for 61 normal control individuals. Affected individuals are indicated by red circles, obligate carriers (parents of affected individuals) by orange triangles, other individuals at risk for inheriting an affected allele by green diamonds, and additional members of family A not at risk by open hexagons. Normal controls are shown as blue triangles. Levels for at-risk individuals (green diamonds in **a**) fall into a bimodal distribution, with one peak ranging from 0.45 to 0.68 U ml⁻¹, consistent with carriers, and the other from 0.90 to 1.17 U ml⁻¹, indistinguishable from the normal distribution shown in **b**. Not shown here, probands E, F and G had VWF-cleaving protease activities of 0.05, 0.04 and 0.09 U ml⁻¹, respectively, and no detectable inhibitor. The parents of proband E had VWF-cleaving protease levels of 0.49 and 0.63 U ml⁻¹ and the parents of proband G had levels of 0.53 and 0.68 U ml⁻¹.

genome scan (families E–G, Table 1). The 12 mutations identified account for all but one of the 15 disease alleles expected in this set of patients (Table 1 and Fig. 5, and see Supplementary Information Figs 6 and 7). No recurrent mutation was observed, except in family A (Fig. 5). All three affected individuals in this family are homozygous for the same mutation carried on the same extended haplotype (data not shown), suggesting recent common ancestry. Although there is no known consanguinity, the parents of affected individuals AIII-2, AIII-3 and AIII-8 are all from the same small village where the families have lived for several generations.

Two of the 12 identified mutations result in frameshifts (a 26-base pair (bp) deletion in exon 19 (Fig. 5) and a single T insertion in exon 27). A single splice mutation was identified in one family (1584 + 5G → A, family G). This substitution markedly reduces or eliminates utilization of the normal intron 13 splice donor and activates a cryptic donor splice site at +70, resulting in a 23-codon insertion (see Supplementary Information Fig. 6). The remaining nine mutations all result in nonconservative amino-acid substitutions (Table 1) and all occur at positions that are perfectly conserved between the human and murine genes. H96D in the catalytic domain and R398H in the first TSP1 domain both alter residues that are almost perfectly conserved among all known ADAMTS family members. All 12 of the *ADAMTS13* gene mutations described above were excluded as common sequence polymorphisms by screening a large panel of unaffected chromosomes. Many single-nucleotide polymorphisms (SNPs) were also identified,

Table 1 ADAMTS13 mutations in thrombotic thrombocytopenic purpura

Exon	Family	Nucleotide	Amino acid
3	B	286C → G	H96D
3	E	304C → T	R102C
6	E	587C → T	T196I
10	D	1193G → A	R398H
13	C	1582A → G	R528G
13	G	1584+5G → A	Splice
17	A	2074C → T	R692C
19	F	2376–2401Δ	Frameshift
22	B	2851T → G	C951G
24	D	3070T → G	C1024G
26	F	3638G → A	C1213Y
27	C	3769–3770insT	Frameshift

Genomic DNA from patients was used to amplify exons and intron–exon boundaries of *ADAMTS13*. For mutations in families A–D, candidate mutations were confirmed in both parents. All mutations were confirmed as absent from ~180 control normal chromosomes by PCR, allele-specific oligonucleotide hybridization or restriction digest.

although only 7 of 25 result in amino-acid substitutions (see Supplementary Information Table 2).

The spectrum of *ADAMTS13* mutations observed here is notable for the relative paucity of obvious null alleles. In addition, the single splice mutation and two frameshift mutations all occur in *trans* with a missense mutation on the other allele. These data suggest that complete deficiency of *ADAMTS13* may be lethal, consistent with the low levels of residual VWF-cleaving protease activity observed in all ten deficient patients studied here (0.02–0.09 U ml⁻¹).

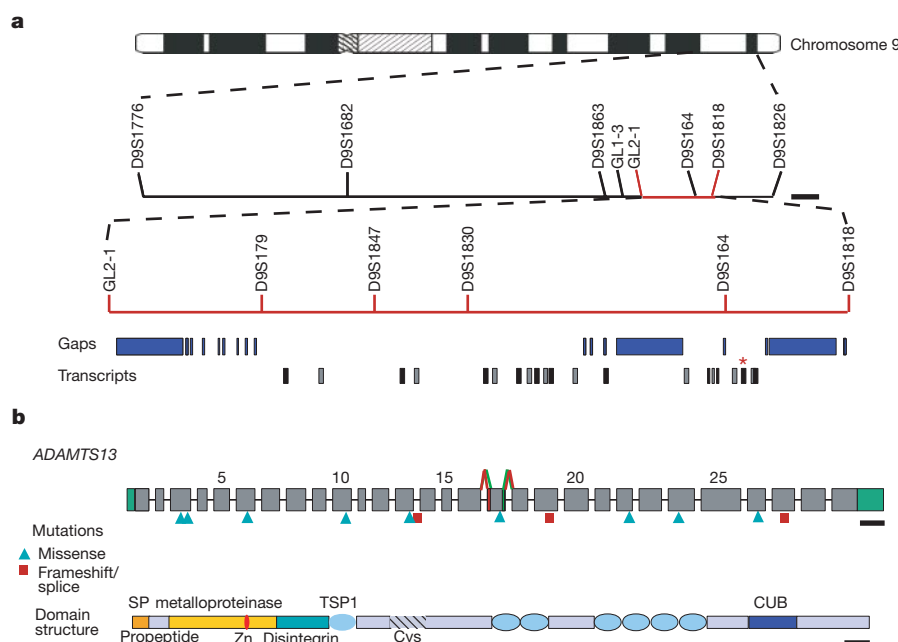


Figure 3 Identification of the *ADAMTS13* gene. **a**, Physical map of chromosome 9 in the interval surrounding marker D9S164. The 2.3-Mb nonrecombinant interval identified in Fig. 1 is shown in red. Sequence gaps in the public genomic draft assembly are denoted by blue rectangles. Transcripts localized to this interval are depicted by alternating black and grey bars. The predicted gene *C9ORF8* is indicated with a red asterisk and is located ~32 kb telomeric to D9S164, on the basis of the current draft human genome sequence¹⁷. The reference bar represents 1 Mb. **b**, Intron–exon and domain structures of *ADAMTS13*. The coding region is illustrated in grey and 5' and 3' untranslated regions in green. Intron sizes are not drawn to scale. Exon 1 of *C9ORF8* overlaps with a cluster of EST sequences (UniGene cluster Hs.149184), initially interpreted as predicting a large 5' untranslated region. A segment of putative *C9ORF8* coding sequence was used to identify two partial cDNA clones from human fetal liver, which were extended in both the 5' and 3' directions by RT-PCR and RACE. The assembled cDNA sequence corrected an error in the predicted boundaries of *C9ORF8* exon 2, resulting in a continuous open reading frame including two exons upstream of the 5' EST cluster, three new exons within

the predicted intron 10 of *C9ORF8* and six additional downstream exons encompassing a second hypothetical gene in this region, *DKFZp434C2322* (Unigene cluster Hs.131433). Analysis of RT-PCR and cDNA sequences identified an alternatively spliced variant of exon 17 using both alternate donor and acceptor splice sites, as indicated. Mutations are depicted underneath the corresponding exons, with green triangles representing missense mutations and red squares representing frameshift and splice mutations. Scale bar, 200 nucleotides. The predicted domain structure of *ADAMTS13* is shown at the bottom. The predicted signal peptide (predicted using SignalP version 1.1 (<http://www.cbs.dtu.dk/services/SignalP>)) is indicated in orange, the short propeptide in blue, the metalloproteinase domain in yellow, the disintegrin domain in green, and TSP1 domains as light blue ovals. The locations of the zinc-binding catalytic consensus within the metalloproteinase domain and the cysteine-rich region within the spacer domain are indicated. The CUB domain (dark blue) has not been identified in other *ADAMTS* family members. Scale bar, 50 amino acids.

Discussion

The findings reported here identify genetic deficiency of ADAMTS13 as the underlying molecular mechanism responsible for most or all cases of familial TTP. This disorder seems to be identical to the congenital microangiopathic haemolytic anaemia also referred to as Schulman-Upshaw syndrome³² (OMIM number 276850). Although symptomatic anaemia and thrombocytopenia are usually present at birth, two of the patients studied here (proband G and BII-2) did not experience their first episodes of TTP until the ages of 4 and 8 years, respectively, despite clinical disease as neonates in siblings of both individuals. This highly variable age of onset, even within the same family, suggests a major contribution of other genetic and/or environmental modifying factors that remain to be identified. Previous studies have demonstrated antibody inhibitors directed against the VWF-cleaving protease in most adult TTP patients, suggesting an autoimmune aetiology^{7,10}. However, the late onset of this disorder observed in two of the ten patients in our series raises the possibility that a subset of apparently sporadic adult-onset TTP may be due to unrecognized recessive inheritance of *ADAMTS13* mutations.

The nearly perfect correlation of plasma VWF-cleaving protease activity with TTP carrier status that we observed was critical to the success of the genetic linkage analysis reported here. Although these data indicate that the plasma levels of this protease activity are very tightly regulated, to date no clinical phenotype has been detected in previously reported familial TTP carriers, or in any of the *ADAMTS13*-mutation heterozygotes identified here. However, the presence of a subtle phenotype in these individuals, such as increased susceptibility to thrombosis or to one or more of the 'acquired' forms of TTP, cannot be excluded.

Our data provide the first direct evidence of an aetiological role for VWF-cleaving protease activity in the pathogenesis of TTP and identify the enzyme likely to be responsible for this activity: metalloproteinase ADAMTS13. Physiologic cleavage of VWF at the Tyr 842–Met 843 peptide bond produces the characteristic 176K proteolytic fragment observed in normal plasma^{5,6,33} and increased susceptibility to this proteolytic cleavage seems to be

responsible for the loss of large VWF multimers central to the pathophysiology of type 2A von Willebrand disease (VWD)^{31,34}.

Taken together with previous work from our group and others^{4–7,10}, our data strongly suggest that ADAMTS13 is the VWF-cleaving protease itself. However, we cannot exclude the possibility that ADAMTS13 acts on VWF indirectly, perhaps by activating another protease. In either case, our data are consistent with the hypothesis that accumulation of large, hyperactive VWF multimers in the absence of normal proteolytic processing triggers pathologic platelet aggregation and is the direct mechanism responsible for TTP. Alternatively, decreased VWF proteolysis could solely be a marker for the loss of ADAMTS13 activity. ADAMTS13 may also have important biological functions elsewhere in the

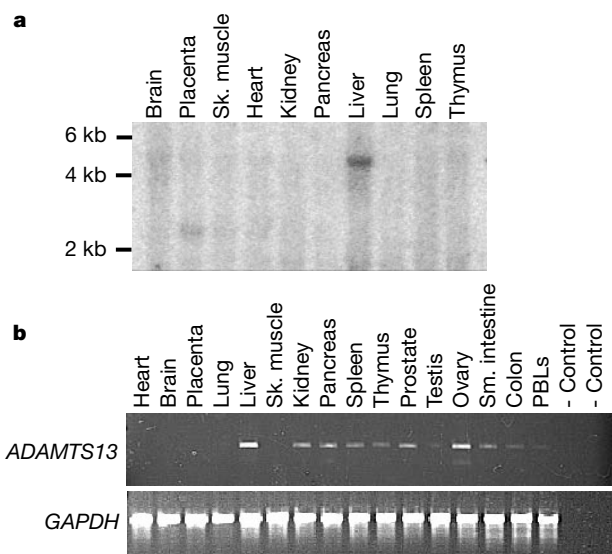


Figure 4 Northern and RT-PCR analysis of *ADAMTS13*. **a**, A northern blot of RNA from multiple human tissues was hybridized with a probe spanning exons 11–13 and part of exon 14. An ~4.7-kb message can be seen specifically in liver and a truncated ~2.3-kb message is faintly visible in placenta. **b**, A panel of cDNAs derived from human tissues was screened by PCR for the presence of exons 10–14. Strong signals were seen in liver and ovary, with weak expression also evident in kidney, pancreas, spleen, thymus, prostate, testis, small (sm.) intestine and peripheral blood leukocytes (PBLs). No expression was detected in heart, brain, placenta, lung or skeletal (sk.) muscle.

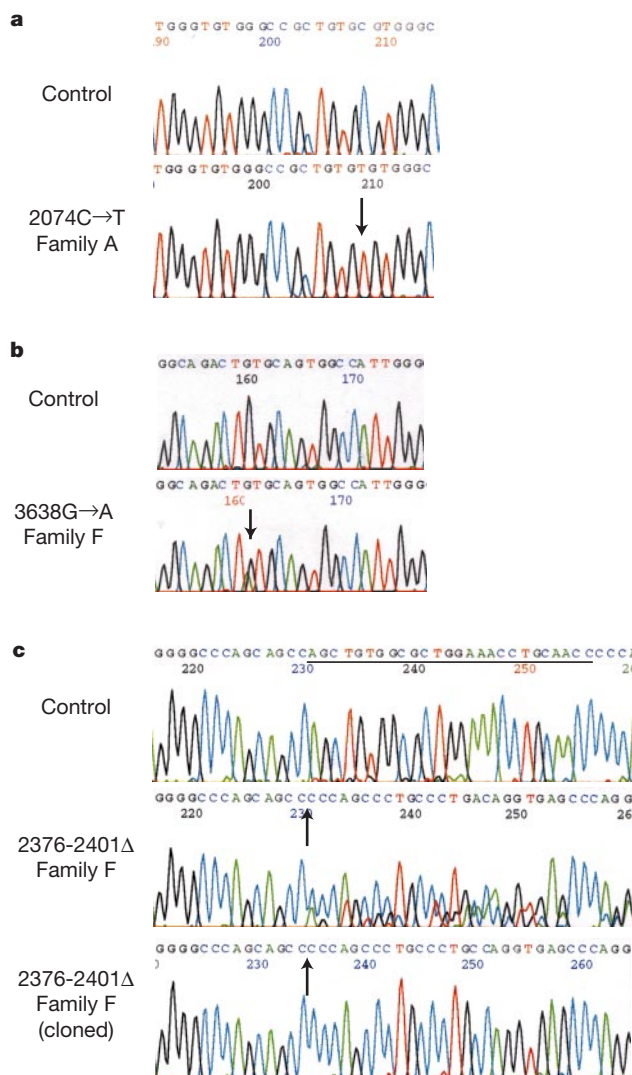


Figure 5 *ADAMTS13* mutations in TTP patients. **a**, Sequence chromatograms of PCR-amplified exon 17 from a control individual and an affected individual from family A. The arrow indicates the position of the 2074C→T (R692C) mutation. The single mutant T peak in the patient sequence chromatogram indicates homozygosity for this mutation. **b**, Sequence chromatograms of PCR-amplified exon 26 identify the heterozygous 3638G→A (C1213Y) mutation (arrowed), accounting for one of the mutant alleles in family F. **c**, Analysis of the second mutant allele in family F. Sequence chromatograms of PCR-amplified exon 19 from a normal control (top panel) and the cloned 2376–2401Δ deletion allele (third panel) are shown. Sequence analysis of total PCR product from the patient (middle panel) demonstrates the superimposed normal and deletion allele sequences, indicating that the patient is heterozygous for this mutation. The underline in the control sequence indicates the 26-bp deletion with the arrows in the other traces indicating the start of the deletion.

coagulation system or in the blood vessel wall, with loss of one or more of these activities providing the direct link to the pathogenesis of TTP. Future studies of ADAMTS13 function, both *in vivo* and *in vitro*, should answer these questions and provide insight into the complex biological processes regulating haemostasis and the maintenance of vascular integrity. Finally, the identification of ADAMTS13 deficiency as the cause of TTP also has major implications for the treatment of this important human disease, including the potential for production of recombinant ADAMTS13 as a safer and more effective replacement for current plasma-exchange therapy.

Note added in proof: Fujikawa *et al.*³⁵ and Gerritsen *et al.*³⁶ have reported purification and sequence analysis of the amino-terminal amino acids of human VWF-cleaving protease. These results confirm the identity of ADAMTS13 as the VWF-cleaving protease and its predicted processing by propeptide convertase cleavage. □

Methods

Subjects

Patients included in this study were referred for evaluation of thrombocytopenia, haemolytic anaemia and schistocytes on blood smear. Proband for the four families (A–D) used in the linkage analysis all had a chronic relapsing course, responded to plasma infusion, and had the disorder as neonates or had a family member with such a disorder as a neonate. All three probands for kindreds E–G also had microangiopathic haemolysis and/or thrombocytopenia noted soon after birth or in the early neonatal period (in the patient or an affected sibling), along with clinical response to plasma infusion.

Plasma samples were obtained from sodium citrate anticoagulated blood by centrifugation and saved at -70°C as previously described⁷. Mononuclear cells were obtained from heparin anticoagulated blood by centrifugation on Ficoll-Hypaque, washed and transformed with Epstein-Barr virus. Informed consent was obtained from all individuals before sample collection (study protocol approved by the institutional review board).

Multimer analysis and VWF-cleaving protease activity

Analysis of VWF multimers was performed as previously described⁷. Ultra large VWF multimers were detected in the plasma of all TTP patients studied, with normal patterns observed in all other family members. For the measurement of VWF-cleaving protease activity, VWF treated with guanidine hydrochloride was used as the substrate. Protease activity was represented by the optical density of the dimer of the 176K fragment generated from the VWF substrate⁷ and was expressed in units per millilitre, with the activity measured in pooled normal control plasma defined as 1 U mL^{-1} . Protease activity is reported as the mean of two measurements for the parents of families E and G. All other samples were measured on at least three occasions, with the mean value reported. Assays for inhibitors of VWF-cleaving protease were performed as described³⁷.

Linkage analysis

A genome-wide linkage screen was performed with 382 polymorphic microsatellite markers spaced an average of 10 cM (panels 1–27 of the ABI Prism Linkage Mapping Set-MD10 (Applied Biosystems)). We amplified 20 ng of genomic DNA with AmpliTaq Gold DNA polymerase (Applied Biosystems). PCR products were run on an ABI Prism 3700 DNA Analyzer and analysed with Genescan version 3.5NT and Genotyper version 3.6NT. Inspection of the pedigrees indicated an autosomal recessive mode of inheritance for TTP in this set of families. The frequency of the disease gene was assumed to be 1 per 10,000 chromosomes in the population. Population frequencies of the marker alleles were estimated from the genotyped individuals. Two-point lod scores were calculated with the program MLINK as implemented in the FASTLINK package version 3.0 (ref. 38), with an autosomal recessive model. A second series of analyses was performed with a codominant model to reflect the lowered enzyme levels of individuals who were assumed to be carriers of the disease gene. For the latter analysis, individuals were classified as affected (those with clinical diagnoses), carriers (those with protease levels in the range of $0.45\text{--}0.68\text{ U mL}^{-1}$) and unaffected (those with protease levels in the range of $0.9\text{--}1.17\text{ U mL}^{-1}$). One individual (AII-2), whose protease level was intermediate between the carrier and unaffected ranges, was classified as phenotype unknown. Penetrance was set at 100% for the recessive model and codominant models. Consistency of the linkage evidence was evaluated by recomputing the lod score with 95% penetrance for the codominant model. Multipoint analyses were performed with the program VITESSE³⁹, using the same two disease models and the five markers at or flanking the maximum two-point lod score (centromere-D9S1682-D9S290-D9S164-D9S1826-D9S158-telomere). Order and distances between markers were determined with the ABI Prism Linkage Mapping Set-MD10 map information. To ensure the multipoint analyses were computationally feasible, the number of founders in pedigree A was reduced and less-informative individuals (II-1, -2, -3 and -13, and III-8 and -9) were pruned from the pedigree.

Sequence analysis

All exons and intron–exon boundaries of the predicted ADAMTS13 gene were amplified from patients' genomic DNA, with the exception of exon 7, which could not be amplified

with multiple primer sets. Intron primers were selected with the Primer3 software package (S. Rozen and H. J. Skaletsky, 1998; code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html) to allow for analysis of exon sequence as well as flanking donor and acceptor splice sites. (See Supplementary Information Table 4 for primer sequences.) We used 100 ng of genomic DNA in a PCR reaction using either Platinum Taq DNA polymerase (Invitrogen), the Expand Long Template DNA polymerase mix (Roche) or the Advantage 2 DNA polymerase mix (Clontech). PCR products were either purified directly from the PCR reaction using the QIAquick PCR purification kit (Qiagen) or gel purified from low-melting agarose (Invitrogen) with the Wizard PCR Preps purification kit (Promega). Total cellular RNA from lymphoblast cell lines was prepared with Trizol (Invitrogen) and RT-PCR performed with the One-Step RT-PCR kit (Invitrogen), according to the manufacturer's instructions. Sequencing reactions were performed by the University of Michigan DNA Sequencing Core. Selected PCR products were subcloned into a pCR-TOPO plasmid (Invitrogen) for further sequence analysis.

Allele-specific oligonucleotide hybridization

Individual exons were amplified from 92 unrelated control individuals. For allele-specific oligonucleotide hybridization, PCR products were spotted onto nitrocellulose membranes by a dot-blot apparatus (Invitrogen). Fifteen-base oligonucleotides corresponding to wild-type or mutant alleles were end-labelled with $\gamma\text{-}^{32}\text{P}\text{-ATP}$ by T4 polynucleotide kinase (New England Biolabs). Hybridization was performed in ExpressHyb solution (Clontech) at 37°C . Blots were washed in $5\times\text{SSPE}$, 0.1% SDS at a temperature determined empirically for each oligonucleotide. Restriction-digested PCR products were analysed on a 3% NuSieve GTG agarose (BioWhittaker), 1% agarose (Invitrogen) gel.

RT-PCR and northern blot analysis

RT-PCR on patient RNA and PCR on cDNA obtained from a Multiple Tissue cDNA panel (Clontech) were performed using primers 5'-CAGTGAACAACCCAGAC-3' and 5'-GGCACCTGTCCCACACCTG-3', which amplify cDNA nucleotides 1,228–1,636 (numbered from the initiator ATG as +1). A FirstChoice Human Northern Blot (Ambion) was screened with a probe generated by random priming using the Rediprime II kit (Amersham) of a PCR product amplified from a human MOLT4 T-cell cDNA library (generated as previously described⁴⁰) with the above primers. Hybridization was performed in ExpressHyb solution (Clontech) according to the manufacturer's specifications. The final wash step was performed in $0.1\times\text{SSC}$, 0.1% SDS at 50°C .

Isolation of cDNA

A human fetal liver cDNA library in $\lambda\text{gt}10$ (Clontech) was screened with the northern probe described above. Two overlapping cDNA clones were obtained, spanning exons 5–14 and 8–20 of the predicted ADAMTS13 cDNA sequence. Phage DNA was purified with a Nucleobond lambda midi kit (Clontech), digested with *EcoRI* (New England Biolabs) and subcloned into pBSII-SK+ (Stratagene). We performed 5' RACE on FirstChoice RACE-Ready human liver cDNA (Ambion). Additional 5' RACE was performed with the 5'/3' RACE kit (Roche) on human liver polyA+ RNA (Ambion). Marathon-Ready human liver cDNA (Clontech) was used for 3' RACE and for an internal RT-PCR, the latter to amplify a product spanning cDNA nucleotides 1,552–2,627. RACE and RT-PCR products were cloned into a pCR-TOPO vector (Invitrogen) and individual clones were subjected to sequence analysis. Nucleotides bridging 5' RACE and cDNA clones (cDNA nucleotides 456–534) were obtained from the C9ORF8 EST cluster (UniGene cluster Hs.149184). Exon–intron boundaries and sequence accuracy were verified against available draft human sequence at <http://www.ncbi.nlm.nih.gov>¹⁷.

Received 30 July; accepted 6 September 2001.

- Cines, D. B., Konkle, B. A. & Furlan, M. Thrombotic thrombocytopenic purpura: a paradigm shift? *Thromb. Haemost.* **84**, 528–535 (2000).
- George, J. N. How I treat patients with thrombotic thrombocytopenic purpura-hemolytic uremic syndrome. *Blood* **96**, 1223–1229 (2000).
- Rock, G. A. *et al.* Comparison of plasma exchange with plasma infusion in the treatment of thrombotic thrombocytopenic purpura. *N. Engl. J. Med.* **325**, 393–397 (1991).
- Moake, J. L. *et al.* Unusually large plasma factor VIII: von Willebrand factor multimers in chronic relapsing thrombotic thrombocytopenic purpura. *N. Engl. J. Med.* **307**, 1432–1435 (1982).
- Tsai, H. M. Physiologic cleavage of von Willebrand factor by a plasma protease is dependent on its conformation and requires calcium ion. *Blood* **87**, 4235–4244 (1996).
- Furlan, M., Robles, R. & Lämmle, B. Partial purification and characterization of a protease from human plasma cleaving von Willebrand factor to fragments produced by *in vivo* proteolysis. *Blood* **87**, 4223–4234 (1996).
- Tsai, H.-M. & Lian, E. C. Y. Antibodies to von Willebrand factor-cleaving protease in acute thrombotic thrombocytopenic purpura. *N. Engl. J. Med.* **339**, 1585–1594 (1998).
- Tsai, H.-M., Rice, L., Sarode, R., Chow, T. W. & Moake, J. L. Antibody inhibitors to von Willebrand factor metalloproteinase and increased binding of von Willebrand factor to platelets in ticlopidine-associated thrombotic thrombocytopenic purpura. *Ann. Intern. Med.* **132**, 794–799 (2000).
- Bennett, C. L. *et al.* Thrombotic thrombocytopenic purpura associated with clopidogrel. *N. Engl. J. Med.* **342**, 1773–1777 (2000).
- Furlan, M. *et al.* von Willebrand factor-cleaving protease in thrombotic thrombocytopenic purpura and the hemolytic-uremic syndrome. *N. Engl. J. Med.* **339**, 1578–1584 (1998).
- Burns, E. R. & Zucker-Franklin, D. Pathologic effects of plasma from patients with thrombotic thrombocytopenic purpura on platelets and cultured vascular endothelial cells. *Blood* **60**, 1030–1037 (1982).
- Siddiqui, F. A. & Lian, E. C. Novel platelet-agglutinating protein from a thrombotic thrombocytopenic purpura plasma. *J. Clin. Invest.* **76**, 1330–1337 (1985).

13. Mitra, D. *et al.* Thrombotic thrombocytopenic purpura and sporadic hemolytic-uremic syndrome plasmas induce apoptosis in restricted lineages of human microvascular endothelial cells. *Blood* **89**, 1224–1234 (1997).
14. Mannucci, P. M. *et al.* Enhanced proteolysis of plasma von Willebrand factor in thrombotic thrombocytopenic purpura and the hemolytic uremic syndrome. *Blood* **74**, 978–983 (1989).
15. Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152–154 (1996).
16. Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L. & Weber, J. L. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63**, 861–869 (1998).
17. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
18. Matsushita, M. & Fujita, T. Ficolins and the lectin complement pathway. *Immunol. Rev.* **180**, 78–85 (2001).
19. Nagase, T. *et al.* Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins *in vitro*. *DNA Res.* **5**, 31–39 (1998).
20. Blobel, C. P. Metalloprotease-disintegrins: links to cell adhesion and cleavage of TNF α and Notch. *Cell* **90**, 589–592 (1997).
21. Kaushal, G. P. & Shah, S. V. The new kids on the block: ADAMTSs, potentially multifunctional metalloproteinases of the ADAM family. *J. Clin. Invest.* **105**, 1335–1337 (2000).
22. Hurskainen, T. L., Hirohata, S., Seldin, M. F. & Apte, S. S. ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of zinc metalloproteases. General features and genomic distribution of the ADAM-TS family. *J. Biol. Chem.* **274**, 25555–25563 (1999).
23. Tang, B. L. ADAMTS: a novel family of extracellular matrix proteases. *Int. J. Biochem. Cell Biol.* **33**, 33–44 (2001).
24. Tang, B. L. & Hong, W. ADAMTS: a novel family of proteases with an ADAM protease domain and thrombospondin 1 repeats. *FEBS Lett.* **445**, 223–225 (1999).
25. Colige, A. *et al.* Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis are caused by mutations in the procollagen I N-proteinase gene. *Am. J. Hum. Genet.* **65**, 308–317 (1999).
26. Shindo, T. *et al.* ADAMTS-1: a metalloproteinase-disintegrin essential for normal growth, fertility, and organ morphology and function. *J. Clin. Invest.* **105**, 1345–1352 (2000).
27. Tortorella, M. D. *et al.* Purification and cloning of aggrecanase-1: a member of the ADAMTS family of proteins. *Science* **284**, 1664–1666 (1999).
28. Fernandes, R. J. *et al.* Procollagen II amino propeptide processing by ADAMTS-3. Insights on dermatosparaxis. *J. Biol. Chem.* **276**, 31502–31509 (2001).
29. Kuno, K. *et al.* Molecular cloning of a gene encoding a new type of metalloproteinase-disintegrin family protein with thrombospondin motifs as an inflammation associated gene. *J. Biol. Chem.* **272**, 556–562 (1997).
30. Bork, P. & Beckmann, G. The CUB domain. A widespread module in developmentally regulated proteins. *J. Mol. Biol.* **231**, 539–545 (1993).
31. Tsai, H.-M. *et al.* Proteolytic cleavage of recombinant type 2A von Willebrand factor mutants R834W and R834Q: Inhibition by doxycycline and by monoclonal antibody VP-1. *Blood* **89**, 1954–1962 (1997).
32. Upshaw, J. D. Jr Congenital deficiency of a factor in normal plasma that reverses microangiopathic hemolysis and thrombocytopenia. *N. Engl. J. Med.* **298**, 1350–1352 (1978).
33. Zimmerman, T. S., Dent, J. A., Ruggeri, Z. M. & Hannini, L. H. Subunit composition of plasma von Willebrand factor. *J. Clin. Invest.* **77**, 947–951 (1986).
34. Lyons, S. E., Bruck, M. E., Bowie, E. J. W. & Ginsburg, D. Impaired intracellular transport produced by a subset of type IIa von Willebrand disease mutations. *J. Biol. Chem.* **267**, 4424–4430 (1992).
35. Fujikawa, K., Suzuki, H., McMullen, B. & Chung, D. Purification of human von Willebrand factor-cleaving protease and its identification as a new member of the metalloproteinase family. *Blood* **98**, 1662–1666 (2001).
36. Gerritsen, H. E., Robles, R., Lämmle, B. & Furlan, M. Partial amino acid sequence of purified von Willebrand factor-cleaving protease. *Blood* **98**, 1654–1661 (2001).
37. Tsai, H. M., Li, A. & Rock, G. Inhibitors of von Willebrand factor-cleaving protease in thrombotic thrombocytopenic purpura. *Clin. Lab.* **47**, 387–392 (2001).
38. Schaffer, A. A., Gupta, S. K., Shriram, K. & Cottingham, R. W. Jr Avoiding recomputation in linkage analysis. *Hum. Hered.* **44**, 225–237 (1994).
39. O'Connell, J. R. & Weeks, D. E. The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. *Nature Genet.* **11**, 402–408 (1995).
40. Ginsburg, D. *et al.* Human von Willebrand factor (vWF): isolation of complementary DNA (cDNA) clones and chromosomal localization. *Science* **228**, 1401–1406 (1985).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank S. J. Weiss for comments on the manuscript; R. L. Nagel, I. I. Sussman, T.-P. Lee, J. Ott and J. E. Sadler for advice and encouragement; and A. Li and S. K. Uniacke for technical assistance. This work was supported in part by grants from the National Institutes of Health to H.-M.T., D.G., T.F. and W.C.N.; D.G. is an investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to D.G. (e-mail: ginsburg@umich.edu). Original cDNA and sequence-tagged site sequences reported here are available from GenBank under accession numbers AF414400, AF414401, G73162, G73163, G73164 and G73165.

Light interference from single atoms and their mirror images

J. Eschner^{*†}, Ch. Raab^{*†}, F. Schmidt-Kaler[†] & R. Blatt[†]

[†] Institut für Experimentalphysik, Universität Innsbruck, Technikerstrasse 25, A-6020 Innsbruck, Austria

^{*} These authors contributed equally to this work

A single atom emitting single photons is a fundamental source of light. But the characteristics of this light depend strongly on the environment of the atom^{1,2}. For example, if an atom is placed between two mirrors, both the total rate and the spectral composition of the spontaneous emission can be modified. Such effects have been observed using various systems: molecules deposited on mirrors³, dye molecules in an optical cavity⁴, an atom beam traversing a two-mirror optical resonator^{5–8}, single atoms traversing a microwave cavity^{9–11} and a single trapped electron¹². A related and equally fundamental phenomenon is the optical interaction between two atoms of the same kind when their separation is comparable to their emission wavelength. In this situation, light emitted by one atom may be reabsorbed by the other, leading to cooperative processes in the emission^{13,14}. Here we observe these phenomena with high visibility by using one or two single atom(s), a collimating lens and a mirror, and by recording the individual photons scattered by the atom(s). Our experiments highlight the intimate connection between one-atom and two-atom effects, and allow their continuous observation using the same apparatus.

Single trapped ions have previously been used for studies of single-atom effects, examples being quantum jumps^{15–17}, antibunching¹⁸, and deterministic entanglement¹⁹. Related experiments have been done with single molecules in solids (see ref. 20 for a review). In our experiment, single Ba⁺ ions are trapped in an ion trap, laser-cooled, and well localized for hours. By retroreflecting the laser-induced fluorescence of a single ion with a lens and a distant mirror, we observe interference fringes with 72% visibility as the mirror distance varies. Simultaneous observation of the light transmitted through the mirror shows that the population of the upper level changes in anticorrelation with the interference fringes, which indicates genuine inhibition and enhancement of the atom's spontaneous emission. When two ions are trapped, they interfere with each other's mirror images, which indicates superradiance and subradiance mediated by the distant mirror. In this case the fringe visibility is 5%. The observed interference contrast, taking into account atomic motion and internal atomic dynamics, is in agreement with the theoretical prediction. The experiment also allows us to study the electromagnetic mode structure around a trapped ion and to determine its position with respect to the mirror with nanometre-resolution.

The experiment is described in Fig. 1. A direct and a retroreflected part of the resonance fluorescence of a single Ba⁺ ion are recorded together on a photomultiplier while the distance between mirror and ion is varied. A scan of fluorescence versus mirror shift is shown in Fig. 2. Interference fringes appear which repeat when the mirror is shifted by half the 493-nm wavelength. The interference contrast (or visibility V) in this example is 72%. The observation shows clearly that light from the ion and from its mirror image, that is, light scattered by the same atom into opposite directions, is coherent and can therefore interfere.

As we will show below, there is the much deeper implication of a back-action on the atom, but first we study the interference signal by writing down an expression for the observed signal intensity, I , assuming an atom at rest and perfect wavefront matching between

the two interfering field components

$$I = \langle |E(t - \tau) + E(t)|^2 \rangle = 2\langle |E(t)|^2 \rangle (1 + \text{Re}[g^{(1)}(\tau)]) \quad (1)$$

Here $\langle \rangle$ denotes time averaging, $E(t)$ is the field component emitted directly into the detector and $\tau = 2(l + d)/c$ (≈ 1.7 ns in Fig. 2) is the delay of the other component which reaches the detector after reflection by the mirror. In the second line of equation (1) we have used the definition of the first-order field correlation function

$$g^{(1)}(\tau) = \frac{\langle E^*(t - \tau)E(t) \rangle}{\langle |E(t)|^2 \rangle} \quad (2)$$

Obviously the last term in equation (1) contains the interference, and its amplitude when d is varied is the visibility,

$$V_{\text{rest}} = |g^{(1)}(\tau)| \quad (3)$$

where the subscript rest indicates that atomic motion has not yet been taken into account.

The emitted field $E(t)$ can be expressed through the atomic polarization which itself is proportional to the atomic operator σ^- (ref. 21):

$$E(t) \propto \sigma^-(t) \quad (4)$$

so we can calculate $g^{(1)}(\tau)$ from the optical Bloch equations (OBEs) that describe the atomic dynamics. For the parameters of Fig. 2 we find (neglecting the back action) $V_{\text{rest}} = 0.98$.

An experimental signature that the interference contrast depends on the internal dynamics of the atom is its dependence on the power of the exciting laser. The visibility is expected to diminish with increasing laser intensity owing to the increasing ratio of inelastic to elastic scattering²¹. In fact, we have observed such a reduction: typically, from $V > 50\%$ it decreases continuously to less than 10% when we change the laser intensity at 493 nm from below saturation to 3-fold saturation. This effect, however, arises only in part from the internal dynamics. It is also caused by the thermal motion of the ion, because laser cooling becomes less efficient when the laser intensity increases. A separation of the two contributions is the scope of future work.

Atomic motion influences the interference in the following way: The ion oscillates along the three axes of the trap, leading to a phase modulation of $E(t)$. A simple calculation shows that a sinusoidal oscillation with amplitude x decreases the interference contrast by a factor $J_0(\eta)$, where J_0 is the 0th-order Bessel function, $\eta = \mathbf{x} \cdot \mathbf{k}_f$ is the modulation index, and $\mathbf{k}_f$ is the fluorescence wave vector. Owing to the ongoing laser cooling, the motional state of the ion is thermal, which corresponds to a gaussian distribution of oscillation amplitudes. This leads to the interference contrast being modified to

$$V_{\text{th}} = V_{\text{rest}} I_0(\eta_{\text{th}}^2) \exp(-\eta_{\text{th}}^2) \quad (5)$$

where I_0 is the 0th-order modified Bessel function and $\eta_{\text{th}} = \sqrt{\langle \eta^2 \rangle}$ is the average (thermal) modulation index.

By inserting into equation (5) the minimum thermal energy of the ion, corresponding to optimum laser cooling parameters, the maximum contrast which could be achieved is 93% for the highest trap frequencies. The reduction to the observed value of $V_{\text{exp}} = 72\%$ arises partly from non-optimal cooling conditions, from fluctuations in the ion–mirror distance caused by acoustic noise, and from imperfect phasefront matching of the two interfering fields. Two sources of phasefront mismatch can be distinguished: distortions on the way to the mirror and back (for example, in the vacuum window) decrease the overall achievable contrast, whereas diffraction at the aperture of lens L1 leads to different phasefront curvatures of the backreflected light and the light emitted directly towards the detector. The latter effect enters into the effective solid angle for which interference takes place,

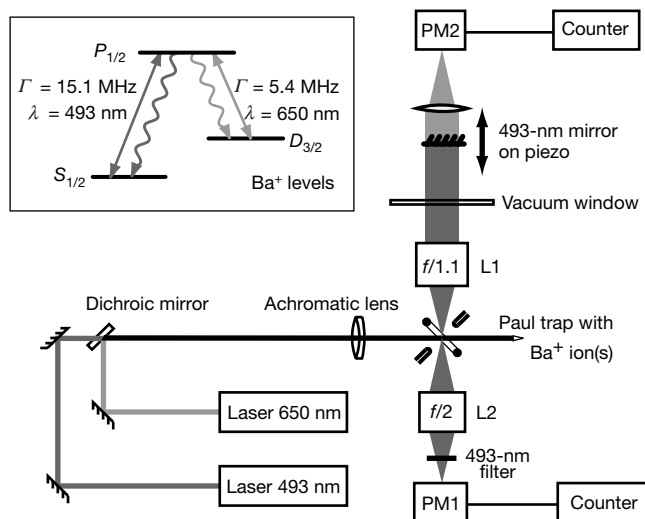


Figure 1 Experimental set-up (main figure) and relevant levels, transition wavelengths, and linewidths of Ba^+ (inset). A single Ba^+ ion is trapped in a Paul trap of 1.4 mm diameter. Its oscillation frequencies ω_z (ω_r) in the trap potential are between 1.2 and 2 (0.6 and 1) MHz. The ion is continuously laser-excited and laser-cooled on its $S_{1/2} \leftrightarrow P_{1/2}$ and $P_{1/2} \leftrightarrow D_{3/2}$ resonance lines. Both lasers have linewidths below 100 kHz; see refs 22 and 23 for details. The laser beams are combined on a dichroic mirror and focused into the trap. Both light fields are linearly polarized and their intensities are set roughly to saturation. The 650-nm laser is tuned close to resonance; the 493-nm laser is red-detuned by about the transition linewidth Γ for Doppler cooling. The precise parameters are determined by fitting a Bloch equation calculation to a scan of the fluorescence intensity versus laser detuning²⁴. A high-quality lens L1 ($f/1.1$, wavefront aberration below $\lambda/5$) at right angles to the excitation beams and 12.5 mm away from the ion collimates the fluorescence light from a solid angle of $\sim 4\%$ into a parallel beam of 21.4 mm diameter. A mirror $l = 25$ cm away retroreflects the 493-nm part of this light while transmitting the 650-nm part. The mirror is angle-tuned for 180° back-reflection with a precision mirror mount and two piezo translators (PZTs). The retroreflected light is focused by L1 to the position of the ion and, together with the light emitted directly into that direction, it is collected with a second lens L2 and recorded with a photomultiplier (PM1). Coarse alignment, that is, superposition of the ion and its mirror image, is controlled visually through L2 while fine adjustment is done by optimizing the signal. The distance between mirror and ion is varied by an amount $d \approx \pm 1 \mu\text{m}$ with another PZT. The 650-nm light transmitted through the mirror is recorded by a second photomultiplier PM2.

but it can be eliminated in the observed interference signal by making the aperture of L2 smaller than that of L1. We checked that an extra aperture stop in front of L2 did not increase the contrast.

In the simple model just presented we have not made use of the particular feature of this experiment that the two interfering light fields are superimposed at the position of the ion rather than on a beam splitter. The description given so far would apply to any beam-splitter model. In contrast, our retroreflecting lens–mirror set-up creates a back-action on the atom which is a fundamentally different effect. Intuitively, this back-action is explained by a modification of the electromagnetic vacuum at the position of the ion. The mirror creates nodes and antinodes in those modes which are collimated by the lens and then retroreflected, among them the modes which are analysed by the detector. The spontaneous emission rate into any of these modes is proportional to the mode intensity at the position of the ion, so we observe reduced or increased fluorescence depending on whether the ion is at a node or antinode; that is, depending on its distance from the mirror. Starting from that idea, the influence of the ion's internal dynamics and of its motion on the interference contrast can also be understood as spectral and spatial broadening,

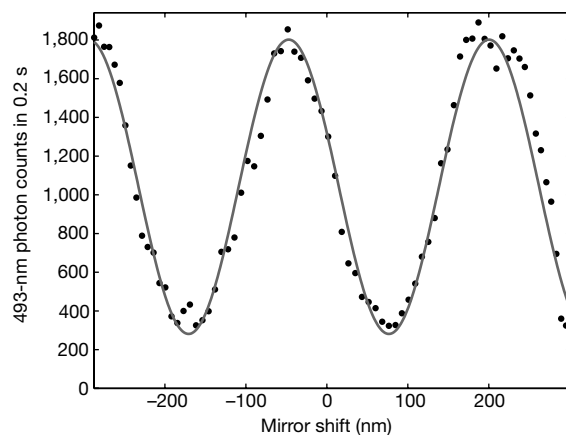


Figure 2 Self-interference in fluorescence of a single atom: photon count rate at PM1 versus mirror displacement (points). The fit (line) accounts for the nonlinear expansion of the PZT with applied voltage. We note that the probability that two photons are interfering is extremely small ($<10^{-5}$), which means that interference does indeed happen in each single emission event.

respectively; the broadening smears out the nodes and antinodes and therefore diminishes the interference contrast.

If some fraction of the total fluorescence is suppressed or enhanced, we also expect the total rate of fluorescence to vary at roughly the same percentage level. An observation of such a variation would verify that a back-action takes place. Therefore we recorded, simultaneously with the interference fringes, the fluorescence at 650 nm, which is transmitted through the mirror (see Fig. 1) and which is directly proportional to the population of the excited ($P_{1/2}$) level of the ion. The result is shown in Fig. 3. The 650-nm fluorescence exhibits a clear $\sim 1\%$ sinusoidal variation anticorrelated with the interference signal, indicating that an interference minimum (maximum) at 493 nm leads to higher (lower) population of the excited state. This shows that the mirror 25 cm away in fact acts on the internal atomic dynamics of the ion by modifying the population of the upper level.

The Methods section explains how this back-action is included into the description of the system by OBEs, the central point being that, in principle, it is indistinguishable whether a photon is emitted directly into the detector or via the mirror. A calculation with the OBEs accordingly modified and with the parameters of Fig. 3 predicts a variation of the total fluorescence by 0.9% if the effective fraction which can be brought to interference, including the various sources of visibility reduction, is set to 1.7%.

The fundamental aspects of the experiment, as described so far, also have fascinating practical implications: we can regard the set-up as a microscope to determine the position of the ion relative to the mirror. The precision of such a measurement is only limited by the noise in the photon-counting signal. It is estimated as follows. With an average count rate (per unit time) i_s and in a measurement interval of duration τ_{int} , we count $I_s = i_s \tau_{\text{int}}$ photons with a poissonian counting error of $\Delta I_s = \sqrt{I_s}$. The maximum slope of the interference signal versus mirror shift is $S = 2\pi A/(\lambda/2)$, where $A = V_{\text{exp}} I_s$ is the interference amplitude in the signal. Thus the error on I_s translates into an error of the position measurement:

$$\Delta x = \frac{\Delta I_s}{S} = \frac{\lambda/2}{2\pi V_{\text{exp}} \sqrt{i_s \tau_{\text{int}}}} \quad (6)$$

which amounts to 1.7 nm in the example of Fig. 2. This means that within a typical measurement time of 0.1 to 1 s the centre position of the ion can be determined more precisely than the extension of its

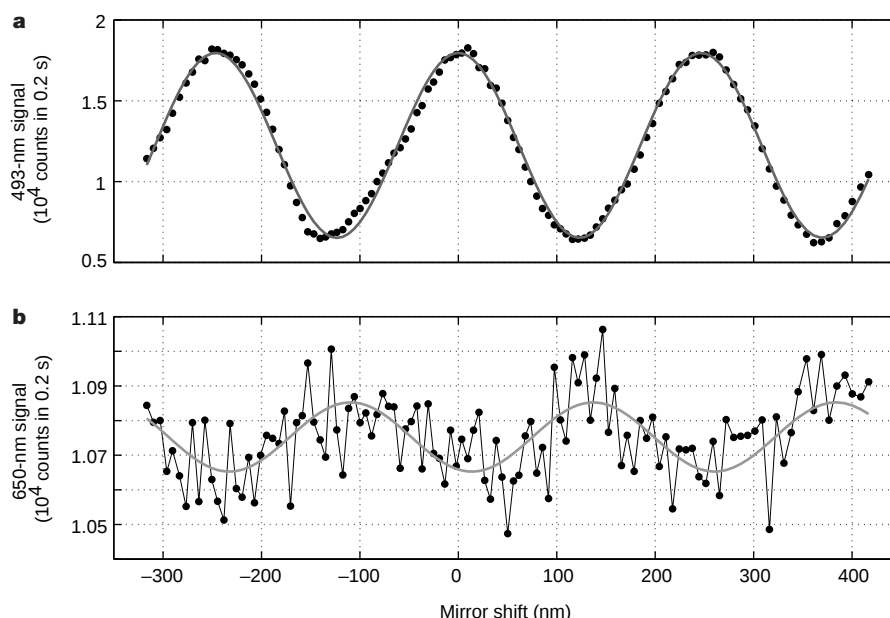


Figure 3 Interference fringes at 493 nm (**a**) and simultaneously recorded fluorescence at 650 nm transmitted through the mirror (**b**). Points are experimental data; bold lines are fits showing sinusoidal oscillations at the same frequency. The visibility of the modulation is 47% (**a**) and 0.9% (**b**). The visibility at 493 nm (**a**) is reduced compared to the

measurement in Fig. 2 because the higher laser intensity that was used here did not permit optimum cooling. Within the experimental error (dominated by the counting noise in the 650-nm signal) the relative phase of the fits is in agreement with anticorrelation, as predicted by the model.

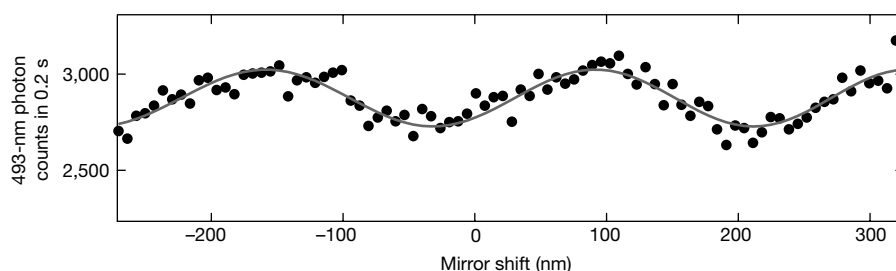


Figure 4 Interference fringes as in Fig. 2 but now with two ions, each interfering with the mirror image of the other. The visibility is about 5%; the main reason for its reduction,

compared to the one-ion experiment, is the strongly driven (micro-) motion of the ions in the Paul trap when their mutual repulsion displaces them from the trap centre.

ground state wave packet in the trap (~ 7 nm). The precision can be increased further by increasing the integration time, which opens up exciting possibilities of measuring and even manipulating the position and motion of the ion on a scale below its position uncertainty. In the same sense, our interference signal reveals spatial variations in the electromagnetic mode structure around the ion on a sub-optical scale; here the resolution is set by the thermal wave packet (~ 35 nm).

We now show that the single-atom phenomena described so far are intimately connected to another fundamental process, namely superradiance and subradiance, which is the cooperative spontaneous emission (or its inhibition) of at least two atoms¹³. With the same set-up as before but with two laser-cooled ions in the trap, we adjust the mirror such that the mirror image of each ion is superimposed with the real image of the other ion. When we scan the mirror we find a result as displayed in Fig. 4. Again, interference fringes appear with the same period as before and with about 5% contrast. However, their interpretation must certainly be different because it is not light from the same atom that interferes, nor is there a back-action of an atom on itself. Instead, the two indistinguishable processes which create the interference are emission by one ion towards the detector and emission by the other towards the mirror, and the two atoms interact with each other.

Inspection of the model for this two-atom case (see the Methods section) reveals the interaction to be reabsorption by one atom of photons emitted by the other. This shows that the observed interference is in fact a signature of subradiance and superradiance. In an earlier experiment¹⁴ the corresponding lifetime modification was studied with two ions spaced by about $1.5 \mu\text{m}$ in a strongly confining trap. In our case, direct interaction between the atoms over their separation of $5 \mu\text{m}$ would not be observable. Instead, the interaction mediated by the mirror over a distance of 50 cm produces a clear and unambiguous effect. $\square$

Methods

The optical Bloch equations (OBEs) derive from a master equation for the atomic density operator ρ , which incorporates the quantum dynamics of the laser-excited atom and the dissipative processes such as spontaneous emission. To include the back-action created by the mirror into the description of the system we have to take into account that the two directions in which a photon can be emitted before it reaches the detector are indistinguishable. This enters into the OBEs in the following way: the term representing dissipation in the master equation is

$$(\dot{\rho})_{\text{diss}} = -\frac{1}{2} \sum_n C_n^\dagger C_n \rho + \rho C_n^\dagger C_n - 2C_n \rho C_n^\dagger \quad (7)$$

where the operators C_n (and their hermitian adjoints $C_n^\dagger$) stand for the different dissipative mechanisms. In particular, spontaneous decay is described by $C = \sqrt{\Gamma} \sigma^-$, the operator σ^-

denoting transitions from the upper to the lower atomic state. This total spontaneous emission can be decomposed, without changing equation (7), into parts $C_1 = C_2 = \sqrt{\Phi\Gamma}\sigma^-$ and $C_3 = \sqrt{(1-2\Phi)\Gamma}\sigma^-$, which stand for spontaneous emission in the direction of the detector, in the opposite direction (towards the mirror), and in the remaining solid angle, respectively. The action of the mirror is now described by replacing C_1 and C_2 in equation (7) by the coherent sum

$$C_{12} = C_1 + e^{-i\omega_l\tau}C_2 = \sqrt{\Phi\Gamma}(1 + e^{-i\omega_l\tau})\sigma^- \quad (8)$$

with laser frequency ω_l , and Φ denoting the effective fraction of the total fluorescence which can be brought to interference, including the contrast reduction by atomic motion, by $g^{(1)}$, and by wavefront distortions. Equation (8) models precisely the situation where the decay that the detector observes is the superposition of the two possible directions, one being delayed by τ . An OBE calculation with this modified dissipation predicts a variation of the total fluorescence with the ion-mirror distance at a level determined by Φ . This is inhibited and enhanced spontaneous emission; it can also be regarded as resulting from reabsorption or stimulated emission induced by the back-reflected photons.

In the two-ion case, the two indistinguishable processes which interfere are emission by one ion towards the detector and emission by the other ion towards the mirror. Using the same tools as above, we model the situation by adding coherently the two decay processes C_{a1} and C_{b2} (and vice versa) where now a and b label the ions and 1 and 2 label the directions of emission. The two new decay operators are given by

$$C_{ab} = C_{a1} + e^{-i\omega_l\tau}C_{b2}; \quad C_{ba} = C_{b1} + e^{-i\omega_l\tau}C_{a2}. \quad (9)$$

When these are introduced into equation (7) (now ρ is the two-atom density matrix), then a term $(\sigma_a^+\sigma_b^- + \sigma_b^+\sigma_a^-)\cos(\omega_l\tau)\rho$ appears in the OBEs, which describes simultaneous emission by one ion and absorption by the other and which is modulated with the distance between the ions via the mirror. This shows that in fact reabsorption (and its inhibition) of the emitted photons goes along with the observed interference. A slightly different viewpoint is that, depending on the phase factor $e^{i\omega_l\tau}$, either the symmetric or the antisymmetric two-atom wave function is preferentially populated, which leads to enhanced or suppressed collective spontaneous emission, respectively. This is superradiance and superradiance as originally described by Dicke¹³.

Received 28 March; accepted 25 July 2001.

- Purcell, E. M. Spontaneous emission probabilities at radio frequencies. *Phys. Rev.* **69**, 681 (1946).
- Milonni, P. W. *The Quantum Vacuum* Ch. 6 (Academic, San Diego, 1994).
- Drexhage, K. H. in *Progress in Optics* (ed. Wolf, E.) Vol. 12, 163–232 (North-Holland, Amsterdam, 1974).
- DeMartini, E., Innocenti, G., Jacobovitz, G. R. & Mataloni, P. Anomalous spontaneous emission time in a microscopic optical cavity. *Phys. Rev. Lett.* **59**, 2955–2958 (1987).
- Jhe, W. *et al.* Suppression of spontaneous decay at optical frequencies: Test of vacuum-field anisotropy in confined space. *Phys. Rev. Lett.* **58**, 666–669 (1987).
- Heinzen, D. J., Childs, J. J., Thomas, J. F. & Feld, M. S. Enhanced and inhibited visible spontaneous emission by atoms in a confocal resonator. *Phys. Rev. Lett.* **58**, 1320–1323 (1987).
- Hood, C. J., Lynn, T. W., Doherty, A. C., Parkins, A. S. & Kimble, H. J. The atom-cavity microscope: Single atoms bound in orbit by single photons. *Science* **287**, 1447–1453 (2000).
- Pinkse, P. W. H., Fischer, T., Maunz, P. & Rempe, G. Trapping an atom with single photons. *Nature* **404**, 365–368 (2000).
- Goy, P., Raimond, J. M., Gross, M. & Haroche, S. Observation of cavity-enhanced single-atom spontaneous emission. *Phys. Rev. Lett.* **50**, 1903–1906 (1983).
- Hulet, R. G., Hilfer, E. S. & Kleppner, D. Inhibited spontaneous emission by a Rydberg atom. *Phys. Rev. Lett.* **55**, 2137–2140 (1985).
- Rempe, G., Walther, H. & Klein, N. Observation of quantum collapse and revival in a one-atom maser. *Phys. Rev. Lett.* **58**, 353–356 (1987).
- Gabriele, G. & Dehmelt, H. G. Observation of inhibited spontaneous emission. *Phys. Rev. Lett.* **55**, 67–70 (1985).
- Dicke, R. H. Coherence in spontaneous radiation processes. *Phys. Rev.* **93**, 99–110 (1954).
- DeVoe, R. G. & Brewer, R. G. Observation of superradiant and subradiant spontaneous emission of two trapped ions. *Phys. Rev. Lett.* **76**, 2049–2052 (1996).
- Nagourney, W., Sandberg, J. & Dehmelt, H. Shelved optical electron amplifier: Observation of quantum jumps. *Phys. Rev. Lett.* **56**, 2797–2799 (1986).
- Sauter, Th., Neuhauser, W., Blatt, R. & Toschek, P. E. Observation of quantum jumps. *Phys. Rev. Lett.* **57**, 1696–1698 (1986).
- Bergquist, J. C., Hulet, R. G., Itano, W. M. & Wineland, D. J. Observation of quantum jumps in a single atom. *Phys. Rev. Lett.* **57**, 1699–1702 (1986).
- Diedrich, F. & Walther, H. Nonclassical radiation of a single stored ion. *Phys. Rev. Lett.* **58**, 203–206 (1987).
- Sackett, C. A. *et al.* Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
- Moerner, W. E. in *Atomic Physics 14* (eds Wineland, D. J., Wieman, C. E. & Smith, S. J.) AIP Conf. Proc. Vol. 323, 467–486 (AIP Press, New York, 1995).
- Loudon, R. *The Quantum Theory of Light* (Oxford Univ. Press, Oxford, 1983).
- Raab, C. *et al.* Diode laser spectrometry at 493 nm for single trapped Ba^+ ions. *Appl. Phys. B* **67**, 683–688 (1998); *Appl. Phys. B* **69**, 253 (1999).
- Raab, C. *et al.* Motional sidebands and direct measurement of the cooling rate in the resonance fluorescence of a single trapped ion. *Phys. Rev. Lett.* **85**, 538–541 (2000).
- Schubert, M., Siemers, I., Blatt, R., Neuhauser, W. & Toschek, P. E. Transient internal dynamics of a multilevel ion. *Phys. Rev. A* **52**, 2994–3006 (1995).

Acknowledgements

We thank P. Zoller, D. Leibfried and G. Morigi for helpful discussions. We gratefully acknowledge support by the European Commission (TMR network QSTRUCT), by the Austrian Science Fund (FWF), and by the Institut für Quanteninformation GmbH.

Correspondence and requests for materials should be addressed to J.E. (e-mail: juergen.eschner@uibk.ac.at).

Bose–Einstein condensation on a microelectronic chip

W. Hänsel, P. Hommelhoff, T. W. Hänsch & J. Reichel

Max-Planck-Institut für Quantenoptik und Sektion Physik der Ludwig-Maximilians-Universität, Schellingstr. 4, D-80799 München, Germany

Although Bose–Einstein condensates^{1–3} of ultracold atoms have been experimentally realizable for several years, their formation and manipulation still impose considerable technical challenges. An all-optical technique⁴ that enables faster production of Bose–Einstein condensates was recently reported. Here we demonstrate that the formation of a condensate can be greatly simplified using a microscopic magnetic trap on a chip⁵. We achieve Bose–Einstein condensation inside the single vapour cell of a magneto-optical trap in as little as 700 ms—more than a factor of ten faster than typical experiments, and a factor of three faster than the all-optical technique⁴. A coherent matter wave is emitted normal to the chip surface when the trapped atoms are released into free fall; alternatively, we couple the condensate into an ‘atomic conveyor belt’⁶, which is used to transport the condensed cloud non-destructively over a macroscopic distance parallel to the chip surface. The possibility of manipulating laser-like coherent matter waves with such an integrated atom-optical system holds promise for applications in interferometry, holography, microscopy, atom lithography and quantum information processing⁷.

Some of the advantages of microscopic magnetic traps on a chip have been pointed out before. Modest electric currents can produce large magnetic field gradients and curvatures in close proximity to a planar arrangement of wires⁸. In an experiment which was realized simultaneously with the results reported here, microfabricated parallel conductors were used in a last stage of evaporative cooling to achieve Bose–Einstein condensation (BEC)⁹. Lithographic fabrication techniques now make it possible to integrate even complex systems of many microscopic traps, waveguides^{10,11}, and other atom-optical devices^{12–14} on a single ‘atom chip’.

The use of such microtraps for BEC appears, in hindsight, only natural, as quantum-mechanical phenomena tend to be more readily observable on a smaller scale. A tight trap permits fast adiabatic changes of the confining potential, and it becomes easy to magnetically compress a trapped atom cloud so that elastic collision times of the order of milliseconds are reached even with just a few million trapped atoms. The resulting fast thermalization makes it possible to drastically shorten the time for radio-frequency-assisted evaporative cooling¹⁵. It is also advantageous that a tight magnetic confinement positions the atom cloud near the centre of the magnetic trap, despite the pull of gravity, so that a rather uniform evaporation is achieved throughout the evaporation process. Collisions with background gas atoms become less important during such a fast cooling cycle, so that the previously very stringent requirements on the vacuum may be greatly relaxed.

Figure 1a shows the chip that is used in our BEC experiments. It features 50- μm -wide conductors, which reproducibly support continuous currents in excess of 3 A. The chip was fabricated

using thin-film hybrid technology, a standard microelectronics process¹⁴. Although a variety of fabrication techniques can be used (several groups have developed custom processes^{16,25}), a standard process has the advantage of being available to non-microtechnology laboratories.

In our chip traps, the trapping potential is produced by superposing the field of the lithographic conductors with a homogeneous, external bias field $\mathbf{B}_0$. A Ioffe–Pritchard potential (non-zero field in the minimum, and quadratic dependence close to this minimum) is created using the ‘Z’-shaped conductor shown in yellow in Fig. 1a; the central part of this conductor has a length of 1.95 mm. Figure 1b shows the magnetic potentials created by this wire with a current of $I_0 = 2$ A for two different values of the external bias field, $\mathbf{B}_0 = B_x \mathbf{e}_x + B_y \mathbf{e}_y$ (where $\mathbf{e}_x$ and $\mathbf{e}_y$ are unit vectors): ($B_x = 0$, $B_y = 8$ G) and ($B_x = 1.9$ G, $B_y = 55$ G), corresponding to the two extreme values used in the experiment. With increasing bias field, the trap centre moves closer to the surface, from $z_0 = 445$ μm to $z_0 = 70$ μm . The potentials vary quadratically with the distance from the trap centre for very small distances. In the main part of the trap, the transverse dependence can be approximately characterized by the gradient of the wire field at z_0 . This gradient rises by a factor of 37—from 200 G cm^{-1} to 7,300 G cm^{-1} —for the two potentials in the figure. The central part of the longitudinal potential, on the other hand, changes very little while its steep walls grow higher with trap compression.

The chip that creates these potentials is mounted face down in a glass cell of inner dimensions $30 \times 30 \times 110$ mm; this cell is part of a simple vacuum system pumped by a 25 l s^{-1} ion pump and a small titanium sublimator (Fig. 2). The magnetic trap lifetime is up to 5 s, which indicates a background pressure in the 10^{-9} mbar

range. Thermal rubidium vapour is produced from a rubidium dispenser¹⁷. We operate it at low, constant, current, so that the rubidium partial pressure remains low and reduces the magnetic trap lifetime by less than 20%. During the magneto-optical trap (MOT) loading phase, a 30-W halogen reflector bulb is switched on to temporarily increase the rubidium pressure by light-induced desorption¹⁸. The MOT loaded in this way contains about 5×10^6 atoms of ^{87}Rb after a loading time of 3–7 s.

Trap loading is a crucial step in realizing a magnetic microtrap. We use the mirror-MOT technique developed in our group and described in detail in refs 5 and 14. This variant of the well-known MOT provides a sample of laser-cooled atoms at a distance of less than 1 mm from the substrate surface, and enables *in situ* transfer into the magnetic microtrap without requiring an intermediate, macroscopic magnetic trap. In this way we typically load 3×10^6 atoms into the magnetic trap after optical pumping into the $F = 2$, $m_F = 2$ ground state.

The initial magnetic trapping potential is shown in dashed lines in Fig. 1b. This trap has frequencies $\nu_x = 28$ Hz and $\nu_{y,z} = 220$ Hz. The temperature and peak density, measured 200 ms after transfer from the MOT, are ~ 45 μK and $\sim 5 \times 10^{10} \text{ cm}^{-3}$, respectively. Immediately after the transfer, we increase the bias field B_y in 300 ms to a final value of 55 G, leading to the potential shown in solid lines in Fig. 1b. Thus, strong compression occurs in the transverse (y, z) plane, whereas the longitudinal potential undergoes a comparatively small change. The very anisotropic final potential has a transverse curvature of $2.4 \times 10^7 \text{ G cm}^{-2}$ near the centre, leading to a calculated transverse oscillation frequency of $\nu_{y,z} = 6.2$ kHz, while the longitudinal frequency is only $\nu_x = 17$ Hz. Experimentally, we have used two different techniques (direct observation of the oscillation and parametric heating) to measure these frequencies, and found good agreement of calculated and measured frequencies.

Immediately after compression, we initiate forced evaporative cooling with a linear radio-frequency sweep from 30 MHz to 8 MHz. This sweep is typically performed in 900 ms, and reduces the number of atoms to 5×10^5 . From here, we can proceed to BEC by two different routes, both of which involve an adiabatic change in the potential. In the first case, the trap is simply decompressed by reducing $\mathbf{B}_0$ —leading to a very elongated condensate centred at the position C1 in Fig. 1a. In the second approach, the wire currents are

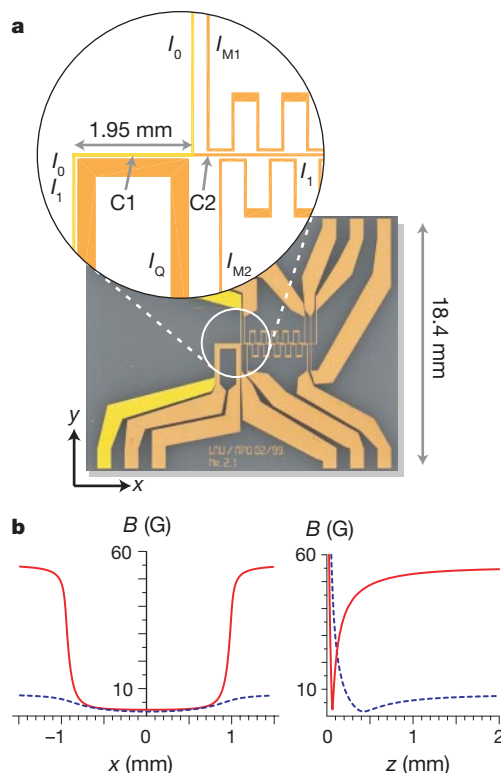


Figure 1 The chip and the magnetic potentials that it creates. **a**, Layout of the lithographic gold wires on the substrate. The inset shows the relevant part of the conductor pattern. I_0 , I_1 , I_{M1} and I_{M2} create the various magnetic potentials for trapping (see **b**) and transport, as described in the main text, I_0 is used only during trap loading (intermediate MOT step⁵). **b**, Potentials created by a wire current $I_0 = 2$ A for two different values of the external bias field, ($B_x = 0$, $B_y = 8$ G) (dashed lines) and ($B_x = 1.9$ G, $B_y = 55$ G) (solid lines).

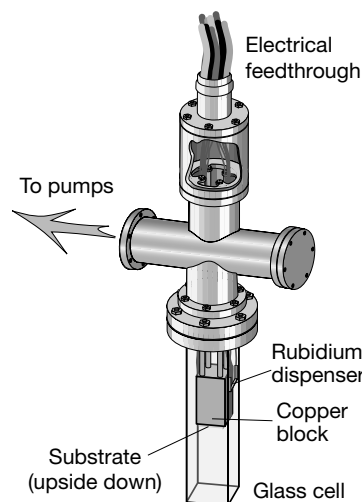


Figure 2 Vacuum system. Because of the efficient evaporation in the chip trap, this very simple set-up is sufficient to achieve Bose–Einstein condensation (BEC). The substrate is mounted facing downwards, so that the atom cloud may be released to expand in free fall (time-of-flight analysis).

also modified to move the trap centre to position C2, where a more spherical condensate arises.

We now describe both cases in more detail. In the first case, the initial radio-frequency ramp is immediately followed by a second, exponential ramp with 800 ms duration. At the end of this ramp, $\sim 7 \times 10^4$ atoms are left at a temperature of $\sim 6 \mu\text{K}$, a density of $5 \times 10^{13} \text{ cm}^{-3}$ and a phase space density in the lower 10^{-2} range. The external bias field is now reduced to ($B_x = 1.2 \text{ G}$, $B_y = 40 \text{ G}$) within 150 ms to decompress the trap. This is done to avoid an excessively high density (which would lead to trap loss by three-body collisions), and to reduce the heating rate, which was measured to be $2.7 \mu\text{K s}^{-1}$ in the compressed trap at $B_y = 55 \text{ G}$, and reduced to $1.1 \mu\text{K s}^{-1}$ at $B_y = 40 \text{ G}$. (The origin of this heating is discussed below.) The decompressed trap has frequencies $\nu_{x,y,z} = (20, 3,900, 3,900) \text{ Hz}$. After a final radio-frequency sweep to $\sim 1.6 \text{ MHz}$ in 300 ms, a condensate appears. Figure 3 shows time-of-flight absorption images after a ballistic expansion time of 21 ms. These images were taken for descending values of the final radio frequency, and show the appearance of the sharp, non-isotropic peak in the momentum distribution which is a key signature of BEC. A bimodal distribution is first observed for a temperature of $T \approx 630 \text{ nK}$, in agreement with the theoretical value of the transition temperature $T_c = 670 \text{ nK}$ (for 11,000 atoms). The condensate lifetime is of the order of 500 ms and can be prolonged to $\sim 1.3 \text{ s}$ when the radio-frequency power is left on at a frequency just above resonance with the condensed atoms. The number of condensate atoms in a distribution well below the transition temperature is typically around 3,000.

The high trap frequencies in the compressed traps currently prevent us from measuring directly the temperature and density in these traps. Instead, we infer these values from analysing the cloud after adiabatic decompression, using a bias field B_y of 19.4 G for the elongated trap, and of 16 G for configuration C2. Detection can then be accomplished by shutting off the wire current, leaving

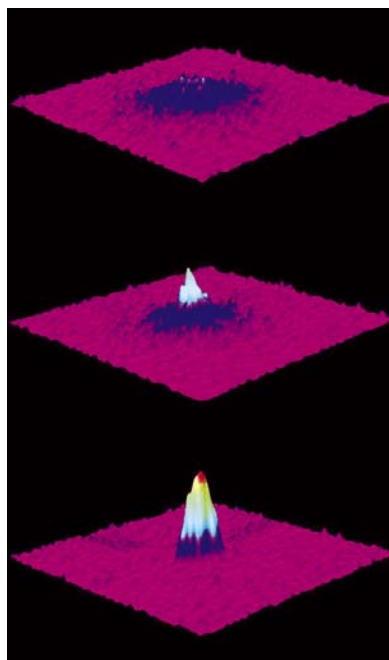


Figure 3 Time-of-flight absorption images of the atom cloud after 21 ms of free expansion. (Profile height and false colours are used to encode the optical density of the cloud.) The images were taken after a final radio-frequency sweep ending at frequencies $\nu_s = \nu_0 + 66 \text{ kHz}$, 44 kHz and 15 kHz (from top to bottom), where ν_0 is the resonance frequency in the centre of the trap. The images show the appearance of the sharp, non-isotropic peak in the momentum distribution which is a key signature of BEC.

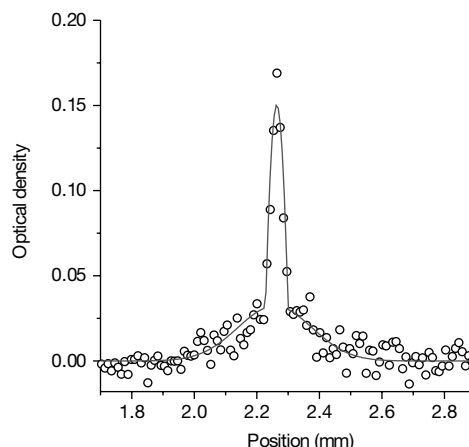


Figure 4 Axial column density profile near the transition temperature, taken in configuration C2. The curve is obtained from a two-dimensional best fit to a bimodal distribution, which corresponds to 1,600 atoms in the condensate and 10,400 atoms in the thermal component at a temperature of $1.7 \mu\text{K}$.

the homogeneous field switched on and tuning the probe laser frequency to the Zeeman-shifted atomic resonance. If the atom cloud is cold enough to remain in the quadratic region of the potential, the temperature in the compressed trap can be calculated from the relation $T' = (\nu'_1 \nu'_2 \nu'_3 / (\nu_1 \nu_2 \nu_3))^{1/3} T$, where ν_i indicates a trap frequency along the i -th axis, and ν'_i designates a value after the transformation. The temperatures we cite are inferred in this way, with the exception of the heating rate measurements below, where all values are given for the detection trap, so that they can be compared directly.

Owing to its very elongated shape, the condensate in position C1 is well inside the regime of fluctuating phase. Such 'quasicondensates' are at present receiving much attention^{19,20}. We also reach the phase transition in a more spherical configuration by transferring the cloud to the position marked C2 in Fig. 1a. This is accomplished by ramping the currents and external fields to final values $I_1 = 2 \text{ A}$, $I_{M1} = 1 \text{ A}$, $B_x = 5 \text{ G}$ and $B_y = 40 \text{ G}$. This transformation takes 250 ms, and results in a trap with frequencies $\nu_{x,y,z} = (300, 3,400, 3,500) \text{ Hz}$ —the ratio of transverse and longitudinal frequencies is now only 12:1, in contrast to configuration C1, where it was 200:1. A single additional radio-frequency ramp with a duration of 500 ms and a final frequency of $\sim 1.0 \text{ MHz}$ is enough to achieve condensation in this trap. The number of

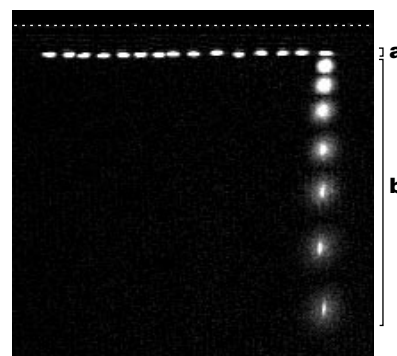


Figure 5 Transport of a BEC on the 'magnetic conveyor belt'. **a**, Superposed absorption images taken at fixed time intervals during transport. The distance between the first and last image is 1.6 mm, the transport time is 100 ms. The line of sight is parallel to the y axis (see Fig. 1); the dotted line marks the edge of the substrate. **b**, Time-of-flight images of the atom cloud after release at the final position, exhibiting the bimodal structure characteristic of a BEC. The maximum expansion time (bottom image) is 19.3 ms.

condensed atoms is typically slightly higher than in configuration C1. A section through a condensed cloud, taken close to the condensation threshold (Fig. 4), reveals the expected bimodal structure with a broad gaussian background of thermal atoms and a sharp peak due to the condensed fraction. When using this potential, we even observe condensation when both radio-frequency ramps are shortened and overlapped with the magnetic field ramp, so that the total evaporation time is as short as 700 ms.

The short evaporation time entails a very fast cycle time of 10 s or less, including MOT loading and detection. Moreover, it relaxes the ultrahigh-vacuum requirements that have been one of the main restrictions of 'traditional' BEC experiments, and thus enables us to use the very simple vacuum system of Fig. 2.

With its trap–substrate distance in the 100- μm range, the microtrapped condensate approaches a room-temperature surface much more closely than do usual condensates. Little is known about condensate–surface interactions at very small distances. Surface-induced heating of trapped atom clouds has been studied theoretically²¹, and is predicted to become important for distances in the 10- μm range. However, there are many other sources that could explain our observed heating rate, such as electrical current noise, mechanical vibrations, and various effects associated with collisions²². We have measured the heating rate after slowly reducing both I_0 and B_y by the same factor α , starting from $I_0 = 2\text{ A}$, $B_y = 40\text{ G}$. This operation reduces both the transverse and longitudinal gradients and curvatures by α while leaving unchanged the position of the cloud centre. We find that the heating rate decreases with α , and reduces to $0.5 \pm 0.3\text{ }\mu\text{K s}^{-1}$ for $\alpha = 0.3$. Although this dependence imposes some restrictions on the nature of the heating mechanism, further measurements are required to fully understand and possibly eliminate it.

A decisive advantage of the chip trap lies in the versatility of the lithographic wire structures. With the wire layout of Fig. 1, many more complicated potentials could be realized. As a first demonstration of these capabilities, we have created a condensate in configuration C2 and transported it over a distance of 1.6 mm along the chip surface using a 'magnetic conveyor belt'⁶. This is done by applying periodically modulated currents I_{M1} and I_{M2} with a relative phase shift of $\pi/2$. Figure 5a shows the cloud during transport over two conveyor periods; images were taken separately and then superposed. Figure 5b shows a series of time-of-flight images in 2.4-ms intervals after release at the final position: the expanding cloud shows the bimodal distribution of a BEC even after transport. This is encouraging for applications such as trapped-atom interferometry^{23,24} with condensates. Because of this robustness, the simplicity of the set-up and the possibilities of chip-based potentials, we expect fruitful results from microchip BEC experiments in the near future. $\square$

Received 27 July; accepted 11 September 2001.

- Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Observation of Bose–Einstein condensation in a dilute atomic vapor. *Science* **269**, 198–201 (1995).
- Davis, K. B. *et al.* Bose–Einstein condensation in a gas of sodium atoms. *Phys. Rev. Lett.* **75**, 3969–1690 (1995).
- Bradley, C. C., Sackett, C. A. & Hulet, R. G. Bose–Einstein condensation of lithium: observation of limited condensate number. *Phys. Rev. Lett.* **78**, 985–989 (1997).
- Barrett, M. D., Sauer, J. A. & Chapman, M. S. All-optical formation of an atomic Bose–Einstein condensate. *Phys. Rev. Lett.* **87**, 010404–1–010404–4 (2001).
- Reichel, J., Hänsel, W. & Hänsch, T. W. Atomic micromanipulation with magnetic surface traps. *Phys. Rev. Lett.* **83**, 3398–3401 (1999).
- Hänsel, W., Reichel, J., Hommelhoff, P. & Hänsch, T. W. Magnetic conveyor belt for transporting and merging trapped atom clouds. *Phys. Rev. Lett.* **86**, 608–611 (2001).
- Calarco, T. *et al.* Quantum gates with neutral atoms: Controlling collisional interactions in time-dependent traps. *Phys. Rev. A* **61**, 022304–1–022304–11 (2000).
- Weinstein, J. D. & Libbrecht, K. G. Microscopic magnetic traps for neutral atoms. *Phys. Rev. A* **52**, 4004–4009 (1995).
- Ott, H., Fortagh, J., Schlottnerbeck, G., Grossmann, A. & Zimmermann, C. Bose–Einstein condensation in a surface microtrap. *Phys. Rev. Lett.* (in the press).
- Müller, D., Anderson, D. Z., Grow, R. J., Schwindt, P. D. D. & Cornell, E. A. Guiding neutral atoms around curves with lithographically patterned current-carrying wires. *Phys. Rev. Lett.* **83**, 5194–5197 (1999).

- Dekker, N. H. *et al.* Guiding neutral atoms on a chip. *Phys. Rev. Lett.* **84**, 1124–1127 (2000).
- Müller, D. *et al.* Waveguide atom beamsplitter for laser-cooled neutral atoms. *Opt. Lett.* **25**, 1382–1384 (2000).
- Cassettari, D., Hessmo, B., Folman, R., Maier, T. & Schmiedmayer, J. Beam splitter for guided atoms. *Phys. Rev. Lett.* **85**, 5483–5487 (2000).
- Reichel, J., Hänsel, W., Hommelhoff, P. & Hänsch, T. W. Applications of integrated magnetic microtraps. *Appl. Phys. B* **72**, 81–89 (2001).
- Ketterle, W. & van Druten, N. J. in *Advances in Atomic, Molecular and Optical Physics* Vol. 37 (eds Bederson, B. & Walther, H.) 181–236 (Academic, San Diego, 1996).
- Drndić, M., Johnson, K. S., Thywissen, J. H., Prentiss, M. & Westervelt, R. M. Micro-electromagnets for atom manipulation. *Appl. Phys. Lett.* **72**, 2906–2908 (1998).
- Fortagh, J., Ott, H., Grossmann, A. & Zimmermann, C. Miniaturized magnetic guide for neutral atoms. *Appl. Phys. B* **70**, 701–708 (2000).
- Anderson, B. P. & Kasevich, M. A. Loading a vapor-cell magneto-optic trap using light-induced atom desorption. *Phys. Rev. A* **63**, 023404–1–023404–4 (2001).
- Petrov, D. S., Shlyapnikov, G. V. & Walraven, J. T. M. Phase-fluctuating 3D Bose–Einstein condensates in elongated traps. *Phys. Rev. Lett.* **87**, 050404–1–050404–4 (2001).
- Dettmer, S. *et al.* Observation of phase fluctuations in Bose–Einstein condensates. *Phys. Rev. Lett.* (in the press); preprint cond-mat/010525 at (<http://xxx.lanl.gov>) (2001).
- Henkel, C., Pötting, S. & Wilkens, M. Loss and heating of particles in small and noisy traps. *Appl. Phys. B* **69**, 379–387 (1999).
- Cornell, E. A., Ensher, J. R. & Wieman, C. E. in *Proc. Int. School of Physics "Enrico Fermi", Course CXL* (eds Inguscio, M., Stringari, S. & Wieman, C.) 15–66 (IOS, Amsterdam, 1999).
- Hinds, E. A., Vale, C. J. & Boshier, M. G. Two-wire waveguide and interferometer for cold atoms. *Phys. Rev. Lett.* **86**, 1462–1465 (2001).
- Hänsel, W., Reichel, J., Hommelhoff, P. & Hänsch, T. W. Trapped-atom interferometer in a magnetic microtrap. *Phys. Rev. A* (in the press).
- Folman, R. *et al.* Controlling cold atoms using nanofabricated surfaces: Atom chips. *Phys. Rev. Lett.* **85**, 5483–5487 (2001).

Acknowledgements

This work was supported in part by the European Union under the IST programme (ACQUIRE project).

Correspondence and requests for materials should be addressed to J.R. (e-mail: jakob.reichel@physik.uni-muenchen.de).

Spatially resolved electronic structure inside and outside the vortex cores of a high-temperature superconductor

V. F. Mitrović*, E. E. Sigmund*, M. Eschrig†, H. N. Bachman*, W. P. Halperin*, A. P. Reyes‡, P. Kuhns‡ & W. G. Moulton‡

* Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA

† Materials Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA

‡ National High Magnetic Field Laboratory, Tallahassee, Florida 32310, USA

Puzzling aspects of high-transition-temperature (high- T_c) superconductors include the prevalence of magnetism in the normal state and the persistence of superconductivity in high magnetic fields. Superconductivity and magnetism generally are thought to be incompatible, based on what is known about conventional superconductors. Recent results¹, however, indicate that antiferromagnetism can appear in the superconducting state of a high- T_c superconductor in the presence of an applied magnetic field. Magnetic fields penetrate a superconductor in the form of quantized flux lines, each of which represents a vortex of supercurrents. Superconductivity is suppressed in the core of the vortex and it has been suggested that antiferromagnetism might develop there². Here we report the results of a high-field nuclear-magnetic-resonance (NMR) imaging experiment^{3–5} in which we spatially resolve the electronic structure of near-optimally doped $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ inside and outside vortex cores. Outside the cores, we find strong antiferromagnetic fluctuations, whereas

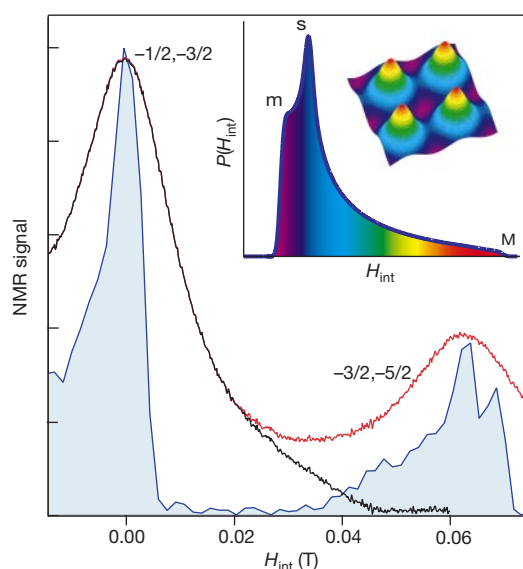


Figure 1 Planar ^{17}O spectra versus the internal magnetic field, H_{int} , showing broadening from vortices. $H_{\text{int}} = \omega/^{17}\gamma - H_0$, where $^{17}\gamma$ is the gyromagnetic ratio for oxygen, and the spectrometer frequency, ω , is set to resonance at the peak frequency for the $-1/2 \leftrightarrow -3/2$ transition in the applied field $H_0 = 13$ T at 11 K. For clarity, only two quadrupolar satellite resonances are shown. In the normal state at 100 K (blue trace) there is a sharp edge that broadens at low temperature (red trace). The spectra are shifted such that internal fields are measured relative to the peak. The black trace represents the $-1/2 \leftrightarrow -3/2$ part of the spectrum obtained by subtraction of the $-3/2 \leftrightarrow -5/2$ transition. The latter was obtained from the convolution of the normal-state spectrum with a broadening function measured independently on the $3/2 \leftrightarrow 1/2$ transition (not shown here). Inset: we have calculated¹⁵ the internal magnetic field profile and the corresponding field distribution function for a 80° vortex lattice at $H_0 = 13$ T, with a penetration depth of $\lambda = 1,500$ Å. The minimum field at the centre of the vortex lattice unit cell (m), the maximum field at the vortex cores (M), and the saddle-point field midway between vortices (s), are shown.

inside we detect electronic states that are rather different from those found in conventional superconductors.

Supercurrents that form a vortex are expected to decay inversely with distance from the vortex core, giving rise to a spatial distribution of internal magnetic fields, $P(H_{\text{int}})$, shown in Fig. 1. The colour coding in the inset to the figure identifies spatial positions in the lattice of vortices with their corresponding location in the field distribution for a vortex lattice. This field distribution will be the same as the NMR spectrum for any nucleus with small intrinsic broadening, as is the case for ^{17}O in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. Its narrow spectrum in the normal state (blue trace) is shown in the figure for two of the four quadrupolar satellite resonances, and compared with the vortex broadened spectrum (red trace) in the superconducting state^{6,7}. We isolate one of the satellite resonances, the $-1/2 \leftrightarrow -3/2$ transition, giving the black trace, corresponding to the expected field distribution shown in the inset. Consequently, our NMR experiment allows a spatially resolved study of the vortex structure. During the NMR experiment, the oxygen nuclei are perturbed and then relax back to thermal equilibrium at the spin–lattice relaxation rate, which is determined by the density of electronic states near the Fermi energy. We have made spin–lattice relaxation rate measurements, Fig. 2, that are also spatially resolved, and we interpret our results in terms of the electronic excitation spectrum both inside and outside the vortex cores. Within the cores, we expect NMR to be sensitive to localized electronic states such as have been observed in conventional superconductors⁸ and recently in high-temperature superconductors^{9,10}. High magnetic fields allow us to use the Zeeman effect to study such electronic excitations. Varying

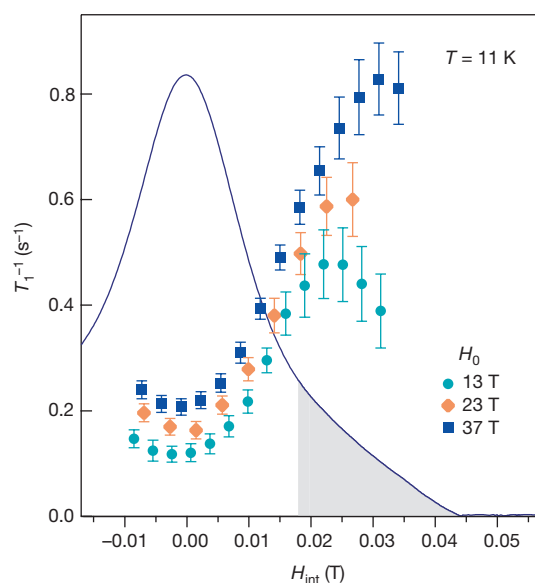


Figure 2 Planar ^{17}O spin–lattice relaxation rate across the vortex lattice NMR spectrum at 11 K. The spectrum is the $-1/2 \leftrightarrow -3/2$ transition at 37 T. The shaded region is the estimated area occupied by the vortex cores at 37 T. We used a superconducting magnet, 13 T; a Bitter-style electromagnet, 23 T; and a hybrid of the two technologies, 37 T. We obtained precise spectra using a field sweep technique, and we measured the relaxation rate using progressive saturation¹⁶.

the field up to 37 T, we sweep through a significant range of corresponding Zeeman energies, $\sim \pm 2.5$ meV, larger than the thermal energy.

Our aligned powder sample of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ is near-optimally doped and $\sim 60\%$ ^{17}O -enriched. The crystal c axis was aligned with the direction of the applied magnetic field. Low-field magnetization data show a sharp transition at $T_c(0) = 92.5$ K. High magnetic fields are needed to increase sensitivity to vortex cores by increasing their fraction of the NMR spectrum. In a field of 37 T, vortices are ~ 80 Å apart, as compared with their core radius given by the size of the electron-pair wavefunction, $\xi = 16$ Å, and so the cores occupy $\sim 15\%$ of both the total volume of the sample and the NMR spectrum. We have shaded the corresponding region in Fig. 2. The high field also suppresses spin diffusion and vortex vibrations, which complicate interpretation of the spin–lattice relaxation at low field⁷.

For high-temperature superconductors, the gap in the excitation spectrum, Δ_k , is believed to have d -wave symmetry, implying that there are four nodes on the Fermi surface where the gap vanishes. At low temperature, the electronic excitations—called quasiparticles—come exclusively from the nodal regions shown in Fig. 3. They can be probed with NMR spin–lattice relaxation, with the rate given by

$$T_1^{-1} \approx \langle |\epsilon - Z + D_i| |\epsilon + Z + D_f| \rangle \quad (1)$$

where ϵ is of the order of $k_B T$, $Z = -(1/2)\gamma_c \hbar H_0$ is the Zeeman energy in the applied field H_0 with gyromagnetic ratio γ_c ; $D_\alpha = \mathbf{v}_{F,\alpha} \cdot \mathbf{p}_s$ is the Doppler shift from the supercurrent momentum $\mathbf{p}_s$; $\mathbf{v}_{F,\alpha}$ is the Fermi velocity, and $\alpha = i, f$ corresponds to initial and final electronic states near the nodes. Depending on the relative magnitude of $k_B T$, Z and D , the dependence of the relaxation rate on temperature and field will show different behaviour for each of three processes, F_1 , F_2 and F_3 defined in Fig. 3. We can vary the first two terms ($k_B T$ and Z) by changing temperature and magnetic field. The Doppler term is proportional to the vortex current, and varies with internal field, H_{int} , across the NMR spectrum. This term is a microscopic analogue of the classical frequency shift of a wave in a moving reference frame, the latter

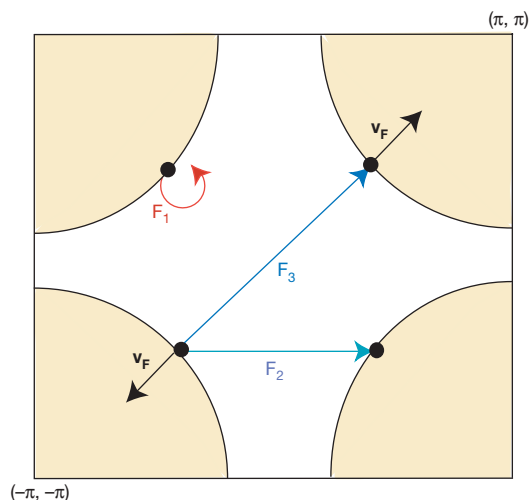


Figure 3 Sketch of the Fermi surface for YBCO and the three processes that contribute to the NMR rate at low temperatures. The shaded regions represent the hole states. The three processes, F_1 , F_2 and F_3 , correspond to the choice of nodes—that is, points where superconductivity is fully suppressed; black dots—for the initial and final quasiparticle states. The spin–lattice relaxation rate is given by $T_1^{-1} \approx \langle N_i(\epsilon)N_f(\epsilon) \rangle$, where $N_\alpha = \sum_{\mathbf{k}} \delta(\epsilon - E_{\mathbf{k},\alpha})$ are the initial ($\alpha = i$) and final ($\alpha = f$) density of states (DOS), respectively. The sum over wavevector $\mathbf{k}$ is restricted to the regions around the nodes, $\mathbf{k}_\alpha$, and the product of the DOS is integrated over energy ϵ in the range $k_B T$ near the Fermi energy, E_F . The excitation energies $E_{\mathbf{k},\alpha}$ for quasiparticles moving in the superflow field with momentum $\mathbf{p}_s$ are Doppler-shifted by an amount $D_\alpha = \mathbf{v}_{F,\alpha} \cdot \mathbf{p}_s$ where the Fermi velocity at node $\mathbf{k}_\alpha$ is given in terms of the normal-state excitation energy $\epsilon_{\mathbf{k}}$ by $\mathbf{v}_{F,\alpha} = \partial \epsilon_{\mathbf{k}} / \partial \mathbf{k}|_{\mathbf{k}=\mathbf{k}_\alpha}$. Note that for process F_3 the relation $D_i = -D_f$ holds, in contrast to, for example, process F_1 where $D_i = D_f$. In sufficiently high magnetic field the Zeeman energy becomes important, and enters initial and final states with different signs ($Z_{i,f} = \pm Z$), leading to $E_{\mathbf{k},\alpha} = \sqrt{\epsilon_{\mathbf{k}}^2 + \Delta_{\mathbf{k}}^2} - Z_\alpha + D_\alpha$. Because the DOS depends linearly on energy near the nodes, the NMR rate is proportional to the product given by equation (1).

being associated with the superconducting condensate.

At the peak of the NMR spectrum, $H_{\text{int}} = 0$ in Fig. 2, we observe that the rate increases as a function of H_0 . This dependence comes from the Zeeman interaction. The Doppler term is small in this region because the supercurrents cancel at the saddle-point. As we increase the applied field, the Zeeman interaction dominates the energy spectrum near the nodes—that is, Z , equivalent to 25 K at 37 T, is larger than $k_B T$ —and from equation (1) the NMR rate rises quadratically with field as $\langle |Z^2 - \epsilon^2| \rangle$.

To the right of the peak in the spectrum towards the vortex core, Fig. 2, we find that the rate increases by a factor of ~ 4 –5. Here the supercurrent momentum increases sufficiently that the Doppler term in equation (1) becomes large compared to both Z and $k_B T$, and it dominates the quasiparticle energy. Our observations show that T_1^{-1} , and consequently the density of electronic states (DOS), depends on distance, r , from the core. In particular, we expect the rate to increase as $T_1^{-1} \approx (p_s)^2 \approx r^{-2}$. The enhancement of T_1^{-1} depends only on r , and should not be dependent on the applied magnetic field other than the Zeeman interaction. This is not to be confused with Volovik's prediction¹¹ that the average DOS is proportional to $\sqrt{H_0}$, which corresponds to $\langle T_1^{-1} \rangle \propto H_0 \ln H_0$ where the average is over the entire spectrum. For conventional superconductors without nodes, the Doppler term has a negligible effect. Our observation of the Doppler shift extending to the saddle-point region shows that, even in high magnetic fields, the superconducting gap has well defined nodes.

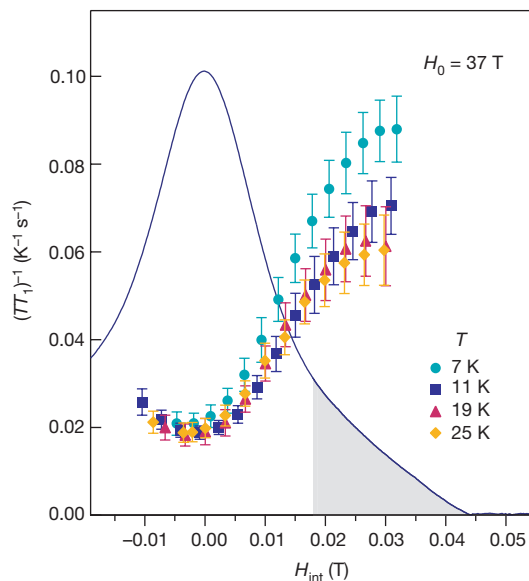


Figure 4 Spin–lattice relaxation rate of planar ^{17}O divided by temperature as a function of internal magnetic field. The shaded region is the area occupied by the vortex cores at 37 T. For a d -wave superconductor in the temperature range where $k_B T$ is much less than $\gamma_e \hbar H_0$, (T_1^{-1}) is expected to be constant. We clearly observe this behaviour near the peak in the NMR spectrum.

Outside the vortex core, we find that the effect of increasing the applied magnetic field from 13 to 37 T is simply to displace the relaxation rate data on a vertical scale in Fig. 2. This is possible if the Doppler term in equation (1) changes sign for initial and final states, just like the Zeeman term, and can only occur if those quasiparticle states are not from the same node (F_2 or F_3). Then, at low temperatures, the rate is proportional to $D^2 + Z^2$ and increases quadratically with field. Otherwise, initial and final states come from the same node (F_1) and the rate is proportional to $D^2 - Z^2$, having the wrong field dependence in the regime $D > Z$. We have calculated the relative strengths of the three possible processes and their dependence on magnetic field. The calculation is performed using the known geometric form factors and the imaginary part of the electronic susceptibility that we extract by fitting the data. To account for our observed field dependence, the spin susceptibility must be strongly peaked near a wavevector corresponding to the F_2 or F_3 process, at least a factor of seven greater than that at zero wavevector. This indicates that strong antiferromagnetic correlations between quasiparticles coexist with superconductivity at low temperatures. Antiferromagnetically correlated quasiparticles have been identified by neutron scattering¹ in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$, a related compound, and are now confirmed in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ from NMR. Lake *et al.*¹ speculated that if there was antiferromagnetism in the vortex core it might polarize the intervening medium and account for their results.

Near the vortex core, the NMR rate decreases after reaching a maximum in an applied field of 13 T (Fig. 2), reflecting a drop-off in p_s . At higher applied fields the rate increases significantly, well beyond its maximum value at 13 T. This increase with field exceeds that measured outside the cores by a factor of four, and is most pronounced at the lowest temperatures, Fig. 4. We attribute this substantial excess rate to vortex core states. When superconductivity is suppressed near surfaces or in the vortex core, localized electronic states, called Andreev bound states, can appear^{8,12}. Their signature is a peak in the density of states at the Fermi energy, called a zero-bias anomaly in tunnelling experiments. Similar behaviour

might be expected in the vortex core of a high-temperature superconductor. If we interpret our relaxation rate in the same way as for quasiparticles outside the cores, following the discussion in the caption to Fig. 3 and equation (1), then both field and temperature dependence are opposite to what is required for the zero-bias anomaly. Consequently, instead of a peak in the density of states there is a mini-gap, $\sim \pm 5$ meV, much sharper than the variation of the DOS outside the vortex core. Previous reports of vortex core states have been based on charge tunnelling in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ (ref. 9) and $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ (ref. 10). Near the Fermi energy ($\sim \pm 5.5$ meV for $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$), a mini-gap structure in the DOS was observed in the vortex core region. This result is inconsistent with theoretical predictions¹³. Furthermore, tunnelling is extremely sensitive to the surface and depends on an understanding of the tunnelling matrix elements¹⁴. In contrast, NMR probes bulk material. And yet both sets of experiments—the present work and refs 9, 10—can have a similar interpretation.

NMR is directly sensitive to the magnetic character of electronic excitations. So an alternative explanation of our results for the vortex core region might involve antiferromagnetism, as predicted by a theory² that unifies superconductivity and antiferromagnetism. Then field-induced antiferromagnetic excitations could strongly enhance the NMR rate, consistent with our observations. In future work, spatially resolved NMR relaxation experiments could be extended to understand the effects of doping on vortex core states, and possibly identify a connection with the spin-pseudogap found in the normal state. □

Received 25 April; accepted 13 August 2001.

1. Lake, B. *et al.* Spins in the vortices of a high-temperature superconductor. *Science* **291**, 1759–1762 (2001).
2. Arovas, D. P., Berlinsky, A. J., Kallin, C. & Zhang, S. C. Superconducting vortex with antiferromagnet core. *Phys. Rev. Lett.* **79**, 2871–2874 (1997).
3. Takigawa, M., Ichioka, M. & Machida, K. Theory of vortex excitation imaging via an NMR relaxation measurement. *Phys. Rev. Lett.* **83**, 3057–3060 (1999).
4. Wortis, R., Berlinsky, A. J. & Kallin, C. Spin-lattice relaxation in the mixed state of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ and Doppler-shifted d -wave quasiparticles. *Phys. Rev. B* **61**, 12342–12351 (2000).
5. Morr, D. K. & Wortis, R. Theory of NMR as a local probe for the electronic structure in the mixed state of the high- T_c cuprates. *Phys. Rev. B* **61**, R882–R885 (2000).
6. Reyes, A. P. *et al.* Vortex melting in polycrystalline $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ from ^{17}O NMR. *Phys. Rev. B* **55**, R14737–R14740 (1997).
7. Curro, N. J., Milling, C., Haase, J. & Slichter, C. P. Local-field dependence of the ^{17}O spin-lattice relaxation and echo decay rates in the mixed state of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. *Phys. Rev. B* **62**, 3473–3479 (2000).
8. Hess, H. F., Robinson, R. B., Dynes, R. C., Valles, J. M. & Waszczak, J. V. Scanning-tunneling-microscope observation of the Abrikosov flux lattice and the density of states near and inside a fluxoid. *Phys. Rev. Lett.* **62**, 214–217 (1989).
9. Maggio-Aprile, I., Renner, Ch., Erb, A., Walker, E. & Fischer, Ø. Direct vortex lattice imaging and tunneling spectroscopy of flux lines on $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. *Phys. Rev. Lett.* **75**, 2754–2757 (1995).
10. Pan, S. H. *et al.* STM studies of the electronic structure of vortex cores in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **85**, 1536–1539 (2000).
11. Volovik, G. E. Superconductivity with lines of gap nodes: density of states in the vortex. *JETP Lett.* **58**, 469–473 (1993).
12. Lesueur, J., Greene, L. H., Feldmann, W. L. & Inam, A. Zero bias anomalies in $\text{YBa}_2\text{Cu}_3\text{O}_7$ tunnel-junctions. *Physica C* **191**, 325–332 (1992).
13. Franz, M. & Tešanović, Z. Self-consistent electronic structure of a $d_{x^2-y^2}$ and a $d_{x^2-y^2} + id_{xy}$ vortex. *Phys. Rev. Lett.* **80**, 4763–4766 (1998).
14. Wu, C., Xiang, T. & Su, Z. Absence of the zero bias peak in vortex tunneling spectra of high-temperature superconductors. *Phys. Rev. B* **62**, 14427–14430 (2000).
15. Brandt, E. H. Precision Ginzburg–Landau solution of ideal vortex lattice for any induction and symmetry. *Phys. Rev. Lett.* **78**, 2208–2211 (1997).
16. Mitrović, V. F., Sigmund, E. E. & Halperin, W. P. Progressive saturation NMR relaxation. *Phys. Rev. B* **64**, 024520-1–024520-7 (2001).

Acknowledgements

We thank J. A. Sauls, J. Moreno, R. Wortis and K. Machida for discussions. This work was supported by the Science and Technology Center for Superconductivity, the Materials Research Center at Northwestern University, and the National High Magnetic Field Laboratory supported by the National Science Foundation and the State of Florida. M.E. acknowledges support from the US Department of Energy, Office of Science.

Correspondence and requests for materials should be addressed to W.P.H. (e-mail: w-halperin@northwestern.edu).

Swollen liquid-crystalline lamellar phase based on extended solid-like sheets

Jean-Christophe P. Gabriel^{*†}, Franck Camerel^{*}, Bruno J. Lemaire[‡], Hervé Desvaux[§], Patrick Davidson[‡] & Patrick Batail^{*}

^{*} Sciences Moléculaires aux Interfaces, FRE 2068 CNRS, 2 rue de Houssinière, BP 32229, F-44322 Nantes Cedex 3, France

[‡] Laboratoire de Physique des Solides, UMR 8502 CNRS, Université Paris Sud, Bâtiment 510, Centre Universitaire, F-91405 Orsay Cedex, France

[§] Service de Chimie Moléculaire, URA 331 CEA/CNRS Saclay, F-91191 Gif-sur-Yvette, France

Ordering particles at the nanometre length scale is a challenging and active research area in materials science. Several approaches have so far been developed, ranging from the manipulation of individual particles^{1,2} to the exploitation of self-assembly in colloids³. Nanometre-scale ordering is well known to appear spontaneously when anisotropic organic moieties form liquid-crystalline phases; this behaviour is also observed for anisotropic mineral nanoparticles^{4,5} resulting in the formation of nematic^{4–7}, smectic⁸ and hexagonal^{9,10} mesophases. Here we describe a lyotropic liquid-crystalline lamellar phase comprising an aqueous dispersion of planar solid-like sheets in which all the atoms involved in a layer are covalently bonded. The spacing of these phosphatoantimonate single layers can be increased 100-fold, resulting in one-dimensional structures whose periodicity can be tuned from 1.5 to 225 nanometres. These highly organized materials can be mechanically or magnetically aligned over large pH and temperature ranges, and this property can be used to measure residual dipolar couplings for the structure determination of biomolecules by liquid-state NMR. We also expect that our approach will result in the discovery of other classes of mineral lyotropic lamellar phases.

Lyotropic liquid-crystalline lamellar phases not only show long-range orientational order but also one-dimensional long-range positional order. They are classically formed (Fig. 1a–c) by self-assembly into layers of organic amphiphilic molecules, block-copolymers or nanorods such as viruses or β -FeOOH colloids^{8,11}. The inorganic layered materials that are used for intercalation purposes¹² display limited swelling (that is, their lamellar period can be increased) upon solvent addition. But these phases rarely show a lamellar period larger than a few nanometres, are hardly fluid and their positional order is lost very quickly upon further swelling. Therefore, such intercalation compounds cannot be considered as forming lamellar liquid-crystalline phases. With large swelling, concentrated aqueous gels of montmorillonite (smectite) clays may still display some long-range orientational ordering, as was recently demonstrated, but no positional long-range order was observed^{13,14}.

In our quest for a lyotropic liquid-crystalline lamellar phase comprised of purely inorganic solid-like sheets, we need to consider low-dimensional solid-state structures built from highly charged covalent single layers that can be swollen. These layers must be chemically stable upon swelling and have the smallest possible counter-ions. This is because electric charges are needed to produce long-range electrostatic repulsions and a small ion size is known to enhance solubility and swelling properties⁵. These criteria are met for the series of solid acids $\text{H}_n\text{M}_n\text{Z}_2\text{O}_{3n+5}$ (with for example $\text{M} = \text{Sb}$, Nb , Ta ; $\text{Z} = \text{P}$, As ; $n = 1, 3$). Of these, we have chosen to study the

[†] Present address: Covalent Materials Inc., 1295 A 67th Street, California 94608, Emeryville, USA.

layered phase, $\text{H}_3\text{Sb}_3\text{P}_2\text{O}_{14}$ (ref. 15; Fig. 1d). This phase is easily made in a three-step synthesis (the detailed, experimental procedure is given in the Supplementary Information), involving first the high-temperature solid-state synthesis of $\text{K}_3\text{Sb}_3\text{P}_2\text{O}_{14}$, then the exchange of potassium cations by protons in concentrated nitric

acid¹⁵, and finally, a thorough rinsing and dialysis of the obtained solid to remove remaining nitrate ions. This leads to homogeneous, transparent, colourless suspensions of $\text{H}_3\text{Sb}_3\text{P}_2\text{O}_{14}$ in water. Their viscoelastic properties depend on the overall mineral volume fraction, ϕ ($\phi = Mc/\rho$; where c is the concentration in mol l^{-1} , M is the molecular mass in g mol^{-1} , and $\rho = 2.4 \text{ g cm}^{-3}$ is the density of $\text{H}_3\text{Sb}_3\text{P}_2\text{O}_{14}$), with dilute suspensions being very fluid.

Observation with the naked eye of a series of test-tubes containing suspensions of decreasing volume fraction between crossed polarizers shows that: (1) samples of high volume fraction ($\phi > 1.78\%$) form birefringent gels (Fig. 2a); (2) for $0.75\% < \phi < 1.78\%$, fluid birefringent suspensions are observed (Fig. 2b); and (3) for $\phi < 0.75\%$ suspensions are biphasic with a well-defined interface between a denser birefringent fluid phase (bottom) and an isotropic, though flow birefringent, phase (top) (Fig. 2c and d). The existence of suspensions that are both very fluid and permanently birefringent unambiguously indicates the thermodynamic nature of the liquid-crystalline order. When examined in natural light the bottom phase shows interference colours ranging from blue to red as ϕ decreases (Fig. 2g). This already shows that there is some positional ordering at the length scale of fractions of a micrometre. This phase behaviour (Fig. 2) is that usually expected for the swelling of a lyotropic lamellar phase: once the maximum swelling is reached, water molecules can no longer be inserted into the interlamellar space and excess water is expelled, leading to phase separation. However, the flow birefringence of the top phase indicates that it contains some anisotropic particles in suspension. This can arise from an equilibrium between the lower and upper phases, which may be viewed as a colloidal 'evaporation/condensation' process, in analogy to a liquid/gas interface.

We established the phase diagram of $\text{H}_3\text{Sb}_3\text{P}_2\text{O}_{14}$ suspensions versus volume fraction and NaCl concentration (Fig. 3). Flocculation appears at large NaCl molarities ($> 0.1 \text{ mol l}^{-1}$). The maximum lamellar period decreases with increasing salt concentration, inducing a shift of the lamellar phase boundary towards a higher volume fraction. Similarly, moving up a vertical line of the phase diagram in the biphasic domain, the volume of the lower birefringent phase decreases, showing that more water is expelled from the interlamellar space as the salt concentration increases. This is strong evidence that electrostatic repulsive interactions are needed to stabilize this lyotropic liquid-crystalline phase. Indeed, such an effect arises when additional salt ions screen the electrostatic forces, reducing their range.

The structure of these aqueous $\text{H}_3\text{Sb}_3\text{P}_2\text{O}_{14}$ suspensions has been further investigated by small-angle X-ray scattering (SAXS) on the

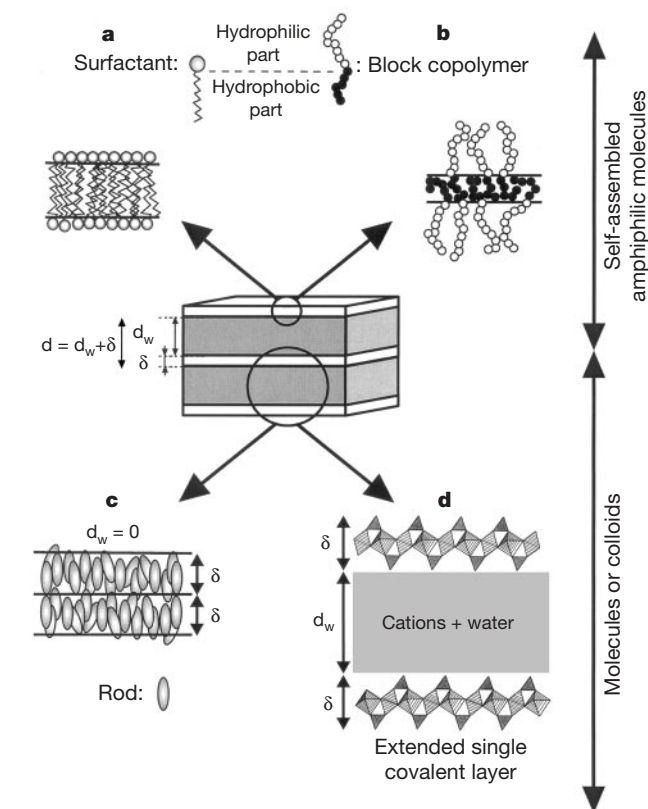


Figure 1 Possible architectures for a liquid crystalline lamellar phase. In a swollen lamellar phase (centre), the lamellae of thickness δ are regularly stacked with a period d and separated by the solvent. The lamellae are usually made from the self-assembly of amphiphilic molecules, such as surfactants (a) or block copolymers (b), or rod-shaped molecules or colloids (c). In contrast, the lamellae considered in this Letter are extended phosphoantimonate covalent single layers made of at least 400 successive unit cells along any direction within the planes (300 nm) (d). The projection of the structure of $\text{H}_3\text{Sb}_3\text{P}_2\text{O}_{14}$ is also shown in d (dark tetrahedra represent PO_4 groups and hatched octahedra represent SbO_6 groups).

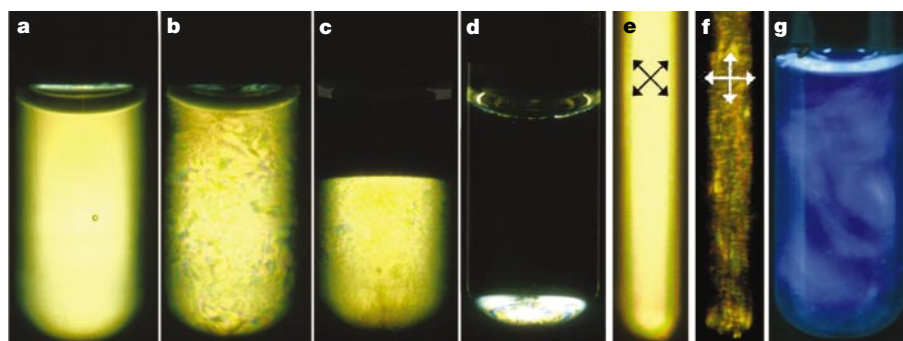


Figure 2 Naked-eye observation of samples. Test-tubes filled with aqueous suspensions of $\text{H}_3\text{Sb}_3\text{P}_2\text{O}_{14}$ single-layers, observed between crossed polarizers (a–e) (the isotropic phase in c and d appears dark). a, in 2 ml of birefringent gel phase (overall volume fraction $\phi = 1.98\%$) the topological defects are so dense that the texture appears homogenous at the scale of this photograph. b, 2 ml of birefringent fluid phase ($\phi = 0.93\%$). c, 2 ml of a biphasic sample ($\phi = 0.65\%$). d, 2 ml of a biphasic sample ($\phi = 0.03\%$).

e and f, Magnetically aligned sample observed in a 5-mm NMR tube that has been immersed for 10 min in a 18.7-T field at 50 °C, in two different orientations compared to the polarizer/analyser system, represented by arrows ($\phi = 0.75\%$). g, Sample iridescence ($\phi = 0.75\%$) observed in natural light is due to light scattering by the $\text{H}_3\text{Sb}_3\text{P}_2\text{O}_{14}$ layers stacked with a period of 225 nm.

ID2 high brilliance beamline at the European Synchrotron Radiation Facility (ESRF)¹⁶. Experiments performed on very dilute isotropic suspensions showed an X-ray scattering intensity, I , that monotonously decreases with scattering vector modulus q ($q = (4\pi \sin \theta)/\lambda$, where 2θ is the scattering angle) as $I \propto q^{-2}$ in all the range of concentrations studied ($0.033\% < \phi < 0.18\%$). This scattering law is typical of two-dimensional planar objects, and its q -range shows that sheets are of considerable lateral spatial extension (at least 300 nm). These experiments prove that the layers remain flat in suspension and do not crumple or fold. At higher volume fractions, typical diffraction patterns of unoriented samples of birefringent suspensions (Fig. 4a) display many sharp peaks (up to 12) that can be indexed to the 001 reflections of a lamellar phase. The small width of these reflections proves that long-range positional ordering does take place, as expected for a lyotropic lamellar phase. The numerous higher-order reflections observed indicate that there is very little positional fluctuation^{17,18} and that the lamellar positional order is very high. Even at maximum swelling, quite a large number (up to 7) of lamellar reflections are still detected. This contrasts with more usual swollen lamellar phases comprised of liquid-like layers, and suggests that this lamellar phase comprised of solid-like sheets has very different elastic constants.

The evolution of the lamellar period d of the birefringent phase upon dilution is shown in Fig. 4b. A first regime, $\phi > 0.75\%$, is observed in which $d \propto 1/\phi$, as expected for the one-dimensional swelling of a pure lamellar phase. For $\phi \approx 0.75\%$, a crossover to a second regime occurs as d levels off and reaches a maximum of about 225 nm. This regime corresponds to the expulsion of excess water in the biphasic domain of the phase diagram. The layer thickness $\delta \approx 1.05 \pm 0.05$ nm can be extrapolated from the first regime, a value that compares very well with that derived from the crystallographic structure, $\delta \approx 1.10 \pm 0.01$ nm. This proves that the layers are perfectly exfoliated. Moreover, as expected, wide-angle X-ray scattering experiments (on samples of volume fractions $\phi > 1.78\%$) showed the existence of fairly thin diffraction lines that arise from the long-range two-dimensional atomic positional order within the covalent layers (Fig. 4a, inset).

The major role of the electrostatic interactions was further

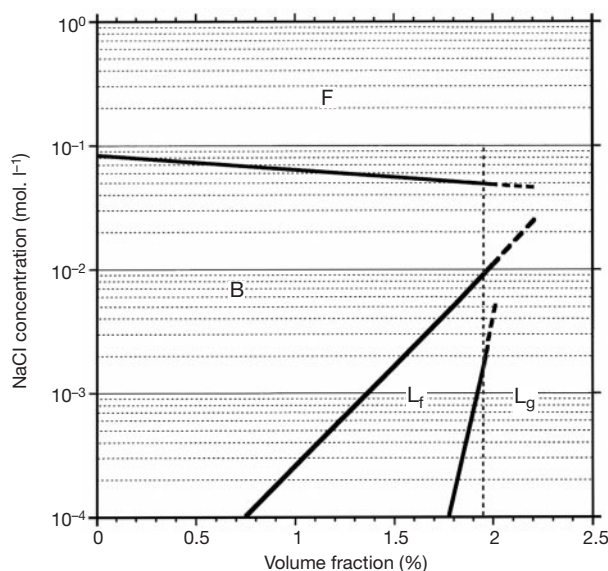


Figure 3 Phase diagram of $\text{H}_2\text{Sb}_3\text{P}_2\text{O}_{14}$ suspensions versus volume fraction and salt concentration. Upon decreasing the volume fraction, the suspensions first form a lamellar gel phase (L_g), then a lamellar fluid phase (L_f) and finally enter a biphasic regime (B). The system flocculates (F) at high salt molarity.

assessed by SAXS experiments (Fig. 4c) in which the lamellar period was measured as a function of salt concentration, namely along the dashed line superimposed on the phase diagram (Fig. 3). d strongly decreases at large ionic strength beyond $[\text{NaCl}] \approx 10^{-2} \text{ mol l}^{-1}$ when the electrostatic interactions are screened, as expected by the DLVO theory of colloidal stability^{19,20}. Thus, at $\phi = 1.9\%$, d falls from 40 nm in pure water down to 23 nm at $[\text{NaCl}] = 0.1 \text{ mol l}^{-1}$.

The lamellar structure of these suspensions was confirmed by X-ray scattering studies of single domains. Aligned samples could easily be produced by mechanical shearing with a Couette cell mounted on the ID2 beamline in both radial and tangential geometries (Fig. 5a), at shear rates, $\dot{\gamma}$, ranging from 30 to $1,000 \text{ s}^{-1}$. At low volume fraction (near the biphasic domain), a strong alignment was induced even at very small shear rate (Fig. 5b). The three possible types of orientation²¹ were observed. However, we can conclude from the difference in scattering intensities that the intuitive c orientation in which the layers orient parallel to the shearing surfaces, is very strongly dominant. The detection of some a and b orientations is very probably due to a few topological

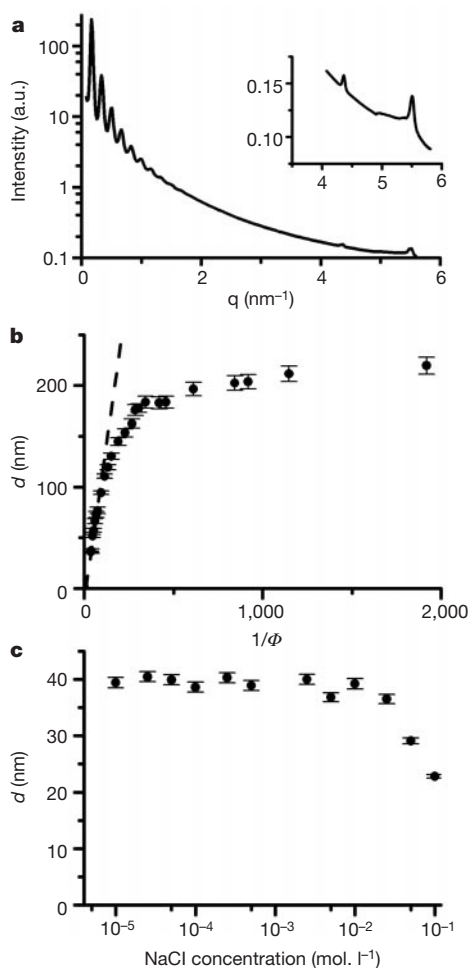


Figure 4 SAXS study of 'powder' samples. **a**, An example of scattered intensity versus scattering vector modulus q , showing 10 orders of reflection due to the lamellar period. Inset, the thin diffraction lines due to the two-dimensional positional order within the covalent layers ($\phi = 2.0\%$). **b**, Variation of the lamellar period, d , with inverse volume fraction (including two data points at very small d extracted from ref. 16). The dashed straight line represents the one-dimensional swelling behaviour of the lamellar phase $d = \delta/\phi$ with $\delta = 1.05$ nm. A crossover from this law to a plateau is observed when entering the biphasic regime. **c**, Variation of the lamellar period d with salt molarity at constant volume fraction ($\phi = 1.9\%$); along the vertical dashed line in Fig. 3.

defects. No relaxation was observed when the shearing was stopped, which proves that the aligned domain is stable (Fig. 5b). Similarly, samples aligned by sucking them into capillary tubes still showed high orientation weeks after their preparation. Comparable results were also obtained with samples of larger volume fractions. However, the degree of alignment of the sheared samples decreased as their concentration increased, probably because of their larger viscoelasticity, which hinders alignment.

Single domains of aqueous suspensions of $\text{H}_3\text{Sb}_3\text{P}_2\text{O}_{14}$ could even be grown in magnetic fields classically used for NMR. Indeed, a SAXS study of these aligned samples showed mosaic spreads as low as 20° (full-width at half-maximum, FWHM) (Fig. 5c) and proved that the layers align their normals parallel to the magnetic field. Also, ^2H NMR spectra obtained from suspensions of $\text{H}_3\text{Sb}_3\text{P}_2\text{O}_{14}$ in deuterated water, D_2O , exhibit a well-resolved residual quadrupolar splitting of the water signal. This property of magnetic alignment, observed in the whole temperature range useful for biomolecules ($4\text{--}50^\circ\text{C}$), could easily be extended at any pH in the range $2.5\text{--}9.5$ by using a dilute base ($c < 0.3\text{ mol l}^{-1}$) with a bulky counter-ion (to avoid flocculation) such as tetramethylammonium hydroxide or tetra(*n*-butyl)ammonium hydroxide (Fig. 2e, f). This medium induces a partial alignment of the dissolved biomolecules, as illustrated in Fig. 6 on an oligosaccharide, leading to residual dipolar couplings, which provide long-range structural restraints that help to determine the structure of the biomolecule by solution-state NMR²². The alignment tensor is significantly different from that induced by vanadium pentoxide suspensions²³, so the simultaneous use of these two mineral liquid crystals is very useful for structure determination²⁴ (see Supplementary Information). By taking advantage of the slow kinetics of the phase separation (a few days), we could even prepare single domains, as observed optically and by NMR, at volume fractions inside the biphasic domain. This procedure could be used to tune the degree of alignment of biomolecules over an even larger range of concentrations. Finally, the possibility of lyophilizing the whole system over a wide range of pH, and the fact that tetramethylammonium hydroxide is commercially available in perdeuterated form, provides a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ -free liquid crystal that seems particularly promising for the structural determination by liquid-state NMR of non-labelled biomolecules.

From a more fundamental perspective, the symmetry properties of this mineral lyotropic lamellar mesophase are very unusual: it has orientational order and exhibits a periodic one-dimensional stacking of covalent solid-like layers. The strong covalent bonds that

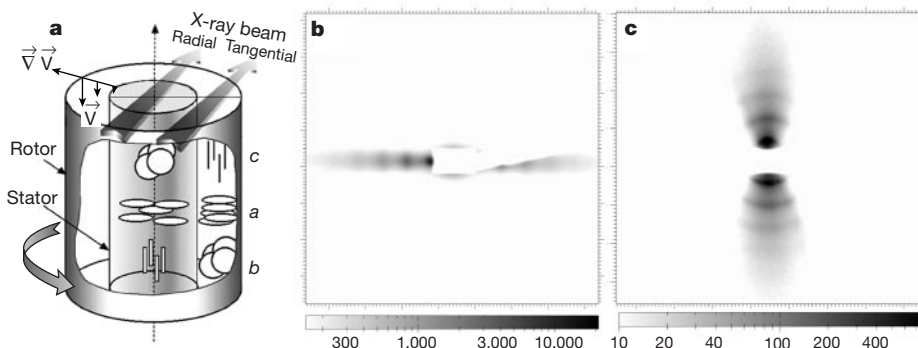


Figure 5 Small-angle X-ray scattering studies of aligned samples. **a**, Scheme of the Couette shear cell illustrating the radial and tangential geometries with respect to the X-ray beam (flat arrows). We also depict within the cell the three different types of lamellar orientations (*a*, *b* and *c*) as seen by the X-rays in both geometries (the layers are sketched as circular disks). **b**, SAXS two-dimensional scattering pattern obtained with a sample aligned in a Couette shear cell, in the tangential geometry, showing that the layers ($d = 175\text{ nm}$) mostly belong to the *c* orientation. **c**, SAXS two-dimensional scattering

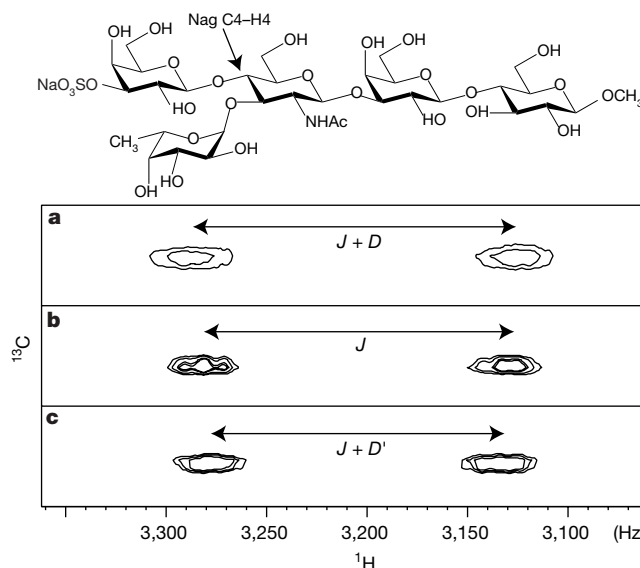


Figure 6 Superposition of three non-decoupled ^{13}C - ^1H NMR sub-spectra²⁹ of the glucosamine (Nag) C4-H4 region of the pentasaccharide containing the Lewis^x motif acquired in a suspension of V_2O_5 (**a**), in isotropic D_2O solution (**b**) and in a suspension of $\text{H}_3\text{Sb}_3\text{P}_2\text{O}_{14}$ in D_2O , $\phi = 0.75\%$ (**c**). In **a** and **c**, the residual dipolar coupling $D = 15.7\text{ Hz}$ and $D' = -5.2\text{ Hz}$ superimposes to the scalar J coupling, confirming the induced alignment of the biomolecule. The different D values obtained using these two media illustrate the different induced alignments (see Supplementary Information).

make up the layers of this mineral lamellar phase are reminiscent of the L_β phase of amphiphilic molecules²⁵. However, one important difference is that the layers of the L_β phase form through the crystallization of self-assembled amphiphilic molecules and melt on heating. In contrast, the mineral lamellar phase, even when swollen, is not sensitive to temperature (up to 70°C where chemical degradation starts taking place within days), owing to the strength of the covalent bonds. This lamellar phase of extended single layers is also an interesting experimental system for testing theoretical predictions describing the properties of periodic stacks of tethered membranes. Indeed, the statistical physics of membranes made up from strong covalent bonds are expected to be very different from those of liquid-like

membranes formed by self-assembly²⁶. The elastic constants and the diffraction peak profiles of single domains of this new phase should be carefully examined and compared to these predictions.

From the point of view of materials chemistry, the approach described here should lead to the discovery of a whole new class of mineral lamellar liquid crystalline phases. Indeed, from current work, we expect that similar phase behaviour will be found not only in other related members of the series $H_3M_3Z_2O_{14}$ (with for example $M = Sb, Nb, Ta; Z = P, As$), but also in phases such as $HM_2Nb_3O_{10}$ ($M = Pb, Ca$), $HTiNbO_5$ or $Zr(HPO_4)_2 \cdot H_2O$ that can all be fully exfoliated using tetra(*n*-butyl)ammonium hydroxide^{27,28}. All of these phases, because of the robustness of their layers, could be used for polymerization of soluble species under tunable two-dimensional confinement conditions. □

Received 8 January; accepted 25 July 2001.

- Eigler, D. M. & Schweizer, E. K. Positioning single atoms with a scanning tunneling microscope. *Nature* **344**, 524–526 (1990).
- Poncharal, P., Wang, Z. L., Ugarte, D. & de Heer, W. A. Electrostatic deflections and electromechanical resonances of carbon nanotubes. *Science* **283**, 1513–1516 (1999).
- Blanco, A. *et al.* Large-scale synthesis of a silicon photonic crystal with a complete three-dimensional bandgap near 1.5 micrometres. *Nature* **405**, 437–440 (2000).
- Davidson, P., Gabriel, J.-C., Levelut, A.-M. & Batail, P. Nematic liquid crystalline mineral polymers. *Adv. Mater.* **5**, 665–668 (1993).
- Gabriel, J.-C. P. & Davidson, P. New trends in colloidal liquid crystals based on mineral moieties. *Adv. Mater.* **12**, 9–20 (2000).
- Kajiwar, K., Donkai, N., Hiragi, Y. & Inagaki, H. Lyotropic mesophase of imogolite 1. Effect of polydispersity on phase diagram. *Makromol. Chem.* **187**, 2883–2894 (1986).
- Nikobakht, B., Wang, Z. L. & El-Sayed, M. A. Self-assembly of gold nanorods. *J. Phys. Chem. B* **104**, 1346–1348 (2000).
- Maeda, Y. & Hachisu, S. Schiller layers in beta-ferric oxyhydroxide sol as an order–disorder phase separating system. *Coll. Surf.* **6**, 1–16 (1983).
- Brown, A. B. D., Ferrero, C., Narayanan, T. & Rennie, A. R. Phase separation and structure in a concentrated colloidal dispersion of uniform plates. *Eur. Phys. J. B* **11**, 481–489 (1999).
- van der Kooij, F. M., Kassapidou, K. & Lekkerkerker, H. N. W. Liquid crystal phase transitions in suspensions of polydisperse plate-like particles. *Nature* **406**, 868–871 (2000).
- Fraden, S. in *Observation, Prediction, and Simulation of Phase Transitions in Complex Fluids* (eds Baus, M., Rull, L. F. & Ryckaert, J. P.) 113–164 (NATO ASI Series, Kluwer Academic, Dordrecht, 1995).
- Whittingham, M. W. & Jacobson, A. J. *Intercalation Chemistry* (Academic, New York, 1982).
- Gabriel, J.-C. P., Sanchez, C. & Davidson, P. Observation of nematic liquid-crystal textures in aqueous gels of smectite clays. *J. Phys. Chem.* **100**, 11139–11143 (1996).
- Mourchid, A., Lécotier, E., Van Damme, H. & Levitz, P. On viscoelastic, birefringent, and swelling properties of laponite clay suspensions: revisited phase diagram. *Langmuir* **14**, 4718–4723 (1998).
- Piffard, Y., Verbaere, A., Lachgar, A., Deniard-Courant, S. & Tournoux, M. The layered phosphoantimonic acid $H_3Sb_3P_2O_{14} \cdot xH_2O$. *Rev. Chim. Miner.* **23**, 766–775 (1986).
- Boescke, P., Diat, O. & Rasmussen, B. High-brilliance beamline at the European Synchrotron Radiation Facility. *Rev. Sci. Instrum.* **66**, 1636–1638 (1995).
- Safinya, C. R. *et al.* Steric interaction in a model multilayer system: a synchrotron study. *Phys. Rev. Lett.* **57**, 2718–2721 (1986).
- Salditt, T. *et al.* Specular and diffuse scattering of highly aligned phospholipid membranes. *Phys. Rev. E* **60**, 7285–7289 (1999).
- Deryagin, B. V. & Landau, L. Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes. *Acta Physicochim. URSS* **14**, 633–662 (1941).
- Verwey, E. J. W. & Overbeek, J. Th. G. *Theory of the Stability of Lyophobic Colloids* (Elsevier, Amsterdam, 1948).
- Safinya, C. R. *et al.* Structure of membrane surfactant and liquid-crystalline smectic lamellar phases under flow. *Science* **261**, 588–591 (1993).
- Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111–1114 (1997).
- Desvaux, H., Gabriel, J.-C. P., Berthault, P. & Camerel, F. First use of a mineral liquid crystal for NMR measurement of residual dipolar couplings of a non-labeled biomolecule. *Angew. Chem.* **40**, 373–376 (2001).
- Ramirez, B. E. & Bax, A. Modulation of the alignment tensor of macromolecules dissolved in a dilute liquid crystalline medium. *J. Am. Chem. Soc.* **120**, 9106–9107 (1998).
- Janiak, M. J., Small, D. M. & Shipley, G. G. Nature of the thermal pretransition of synthetic phospholipids: Dimyristoyl- and Dipalmitoyllecithin. *Biochemistry* **15**, 4575–4580 (1976).
- Toner, J. New phase of matter in lamellar phases of tethered, crystalline membranes. *Phys. Rev. Lett.* **64**, 1741–1744 (1990).
- Fang, M. *et al.* Layer by layer growth and condensation reactions of niobate and titanoniobate thin films. *Chem. Mater.* **11**, 1526–1532 (1999).
- Kaschak, D. M. *et al.* Chemistry on the edge: a microscopic analysis of the intercalation, exfoliation, edge functionalization and monolayer surface tiling reactions of α -zirconium phosphate. *J. Am. Chem. Soc.* **120**, 10887–10894 (1998).
- Kay, L. E., Keifer, P. & Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663–10665 (1992).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank the LURE and ESRF synchrotron radiation facilities for the award of beamtime, C. Bourgaux, P. Panine and T. Narayanan for technical support at D24 (LURE) and ID2 (ESRF), Y. Piffard for supplying us with $1g$ of $H_3Sb_3P_2O_{14}$ powder at the start of this work, C. Chaudemanche and X. Leguevel for their participation in the synthesis of large amounts of $H_3Sb_3P_2O_{14}$, F. Alvarez for pH measurements, S. Grolleau for TGA measurements, P. Berthault and D. Jeannerat for their help in NMR measurements and processing of the spectra, P. Sinaÿ and Y. Zhang for the gift of the pentasaccharide and J. P. Simorre for giving us access to a 18.7 T magnet while we were performing our SAXS experiments at the ESRF. Financial support from the Ministry of Education (PhD fellowship for F.C.), the Ecole Normale Supérieure and the Ecole Nationale des Ponts et Chaussées (PhD fellowships for B.L.), the Région Pays de Loire and the GDR-CNRS 690 FORMES (for D₂O) is gratefully acknowledged.

Correspondence and requests for materials should be addressed to J.-C.P.G. (e-mail: jeang@covalentmaterials.com) or P.D. (e-mail: davidson@lps.u-psud.fr).

High frequency of 'super-cyclones' along the Great Barrier Reef over the past 5,000 years

Jonathan Nott* & Matthew Hayne†

* School of Tropical Environment Studies and Geography, James Cook University, PO Box 6811, Cairns, Queensland, 4870, Australia

† Australian Geological Survey Organisation, GPO Box 378, Canberra, Australian Capital Territory, 2601 Australia

Understanding long-term variability in the occurrence of tropical cyclones that are of extreme intensity is important for determining their role in ecological disturbances^{1–5}, for predicting present and future community vulnerability and economic loss⁶ and for assessing whether changes in the variability of such cyclones are induced by climate change⁷. Our ability to accurately make these assessments has been limited by the short (less than 100 years) instrumented record of cyclone intensity. Here we determine the intensity of prehistoric tropical cyclones over the past 5,000 years from ridges of detrital coral and shell deposited above highest tide and terraces that have been eroded into coarse-grained alluvial fan deposits. These features occur along 1,500 km of the Great Barrier Reef and also the Gulf of Carpentaria, Australia. We infer that the deposits were formed by storms with recurrence intervals of two to three centuries^{8–11}, and we show that the cyclones responsible must have been of extreme intensity (central pressures less than 920 hPa). Our estimate of the frequency of such 'super-cyclones' is an order of magnitude higher than that previously estimated (which was once every several millennia^{12–14}), and is sufficiently high to suggest that the character of rainforests and coral reef communities were probably shaped by these events.

In tropical, intraplate settings that are largely free from the effects of sizeable earthquakes, tsunami, volcanic eruptions and landslides, tropical cyclones rank as one of the main hazards affecting both natural and human communities. In natural communities, disturbance by hazards is one of the most important mechanisms affecting the evolution and diversity of species^{1–5}. Because of the short timescales over which direct and continuous observation of change within ecological communities is possible⁵, it has been difficult to assess the influence of various scales of disturbance, and whether other mechanisms such as recruitment limitation¹⁵ are

significant. The short length of the instrumented record also presents a similar conundrum for human communities. Next to drought, tropical cyclones have the greatest effect on economies and human life of any natural hazards¹⁶. The insurance industry relies on models of future economic loss that use records of cyclone activity extending back usually no more than a century and often less than 50 yr—periods too short from which to make reliable predictions. Cyclones have also been predicted to increase in magnitude and possibly frequency under an enhanced greenhouse climate^{7,17}, but assessing whether any variation is a function of anthropogenic effects requires decoupling recent and future trends from longer-term periodicities. Millennial-scale records of tropical cyclone frequency have been obtained for many areas of the Great Barrier Reef^{8–11} (GBR), but to date there has been no information available on the intensity of these events. With the inclusion of such information it would be possible to accurately assess long-term periodicities and trends in the behaviour of tropical cyclones—particularly those of extreme intensity, referred to here as super-cyclones (extreme category 4 and category 5 storms on the Saffir–Simpson scale).

Storm ridge development occurs when cyclonic waves and surge deposit lithic and biogenic (coral and shell) sediment farther inland than the highest astronomical tide. Most ridges are composed of single storm deposits, which are made up of several facies including storm beach face, berm, crest and washover elements. Beach face and berm facies include porous, clast supported, coarse biogenic shingle deposits that occasionally dip seawards but are usually structureless. Crest facies are horizontally bedded, and are finer-grained than beach-face deposits. Washover facies are bedded, dip landwards up to 15° and sometimes contain imbricated clasts (imbrication is a sedimentary fabric where particles are arranged in an overlapping shingle-like pattern dipping in one direction). Storm deposits are often separated by 'groundsurfaces', which are lenses of pumice pebbles and a weak sooty or earthy palaeosol^{10,11}. An individual storm deposit is generally thought to represent one storm event, on the basis of sedimentary architecture, groundsurfaces and detailed chronologies. Supporting this assumption are observations of single coral shingle storm

ridges—up to 18 km long, 37 m wide and 3.5 m high—being deposited during individual tropical cyclones^{18–21}. The storm deposits analysed in this study occur as multiple, shore-parallel ridges on tectonically stable coasts and islands (Fig. 1a, b). The age of each consecutive deposit/ridge increases progressively with distance from shore (Fig. 1b).

Terraces occur when waves and surge leave erosional scarps. They occur within raised gravel beach deposits, in this instance reworked from coarse-grained alluvial fans. These ridges and terraces are a function of mean storm 'still' water level (storm surge, tide and wave set-up) and wave run-up. (Here, 'surge' is the rise in water level due to lowered atmospheric pressure and surface wind stress; 'set-up' is the rise in water level due to the action of breaking waves; and 'run-up' is the rush of water up against a structure or beach.) The frequency and intensity of past cyclones are determined by geologically dating and topographically surveying these features, then running numerical models that estimate surge and wave heights necessary to reach these levels (see Methods).

The ridge sequences used in this analysis occur at Princess Charlotte Bay (PCB) towards the northern end of the GBR, Fitzroy, Normanby and Curacoa islands within the central GBR, and Lady Elliot island in the south. Wallaby island lies in the Gulf of Carpentaria. A sequence of four erosional terraces occur at Red Cliff (Fig. 1a). Radiocarbon dates from these deposits reveal that the storms responsible have a mean return interval of 177 yr at PCB and 280 yr at Curacoa island over the past 5,000 yr (refs 8, 10, 11), 253 yr at Lady Elliot island over the past 3,200 yr (ref. 9), and approximately 180 yr at Wallaby island over the last 4,100 yr (see Methods). The ridges at Fitzroy and Normanby islands and terraces at Red Cliff are fewer in number and formed over the past few hundred years (Table 1).

Our analyses show that most ridges on Fitzroy, Normanby, Curacoa, Lady Elliot and Wallaby islands and the highest terraces at Red Cliff were emplaced during category 5 cyclones, which probably had central pressures of less than 920 hPa (Fig. 2; Table 2). Even at the 95% uncertainty margin, these cyclones were still category 5 or at least severe category 4 events. The storms responsible for construction of the ridges at PCB appear to have

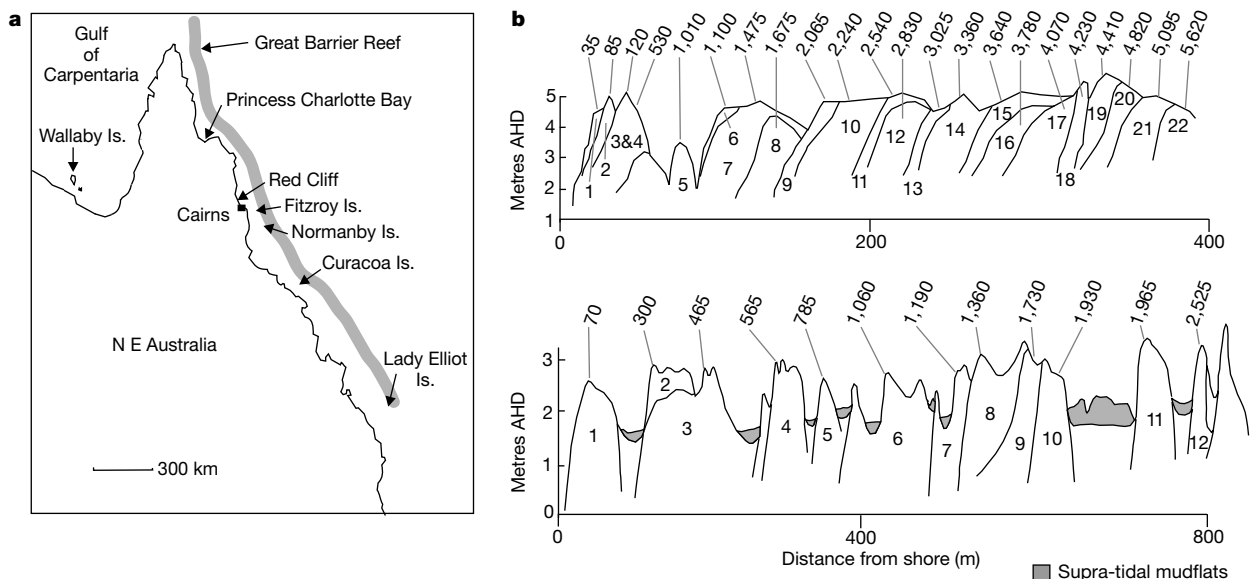


Figure 1 Study sites and storm deposit data. **a**, Location map of study sites. **b**, Stratigraphic relationship of storm deposit/ridges on Curacoa island (top) and Princess Charlotte Bay (bottom). Successive storm deposits are numbered accordingly. Mean

reservoir-corrected radiocarbon age (in yr BP) for each ridge is shown above traces. Note progressive increase in age with distance inland. Age details in Table 1. Cross-sections modified from refs 10 and 11. AHD is Australian Height Datum.

Table 1 Radiocarbon ages for coral shingle storm deposits

Location	Sample code	Radiocarbon age (yr BP)	Calibrated 2 σ age (yr BP)
Wallaby 1	ANU 9229	350 $\pm$ 80	290–480
Wallaby 3	ANU 9230	570 $\pm$ 90	520–650
Wallaby 21	ANU 9233	2,460 $\pm$ 110	2,250–2,640
Wallaby 24	ANU 9234	3,350 $\pm$ 80	3,180–3,460
Wallaby 31	ANU 9236	4,140 $\pm$ 90	3,950–4,270
PCB 1	ANU 8896, 8897	70 $\pm$ 70	0–280
PCB 2	ANU 9037, 8901, 8898	300 $\pm$ 45	290–480
PCB 3	ANU 8899, 8902–8905	465 $\pm$ 30	480–545
PCB 4	ANU 8907–8909, 8911	565 $\pm$ 30	520–650
PCB 5	ANU 8914, 8915	785 $\pm$ 50	660–910
PCB 6	ANU 8916–8919, 8923	1,060 $\pm$ 35	930–1,060
PCB 7	ANU 8922, 8924, 8925	1,190 $\pm$ 35	1,000–1,270
PCB 8	ANU 8926–8930, 8932	1,360 $\pm$ 30	1,230–1,340
PCB 9	ANU 8934, 8936, 8937	1,730 $\pm$ 30	1,550–1,720
PCB 10	ANU 8938–8940	1,930 $\pm$ 60	1,720–1,960
PCB 11	ANU 8942–8945	1,965 $\pm$ 40	1,840–2,050
PCB 12	ANU 8946–8950, 8941	2,525 $\pm$ 50	2,350–2,760
Red Cliff 4	Wk 9613–9616	400 $\pm$ 50	0–170
Fitzroy 1	Wk 7979, 7980	Modern	
Fitzroy 2	Wk 7983, 7984	380 $\pm$ 50	0–148
Normanby 1	Wk 8019, 8017	Modern	
Normanby 2	Wk 8016, 8018	380 $\pm$ 50	0–148
Curacoa 1	ANU 8477, 9932, 9936	35 $\pm$ 35	30–260
Curacoa 2	ANU 9931, 9933, 9934	85 $\pm$ 40	20–270
Curacoa 3	ANU 9897	120 $\pm$ 80	10–310
Curacoa 4	ANU 8479, 8480, 9935	530 $\pm$ 50	500–650
Curacoa 5	ANU 8482–8486	1,010 $\pm$ 50	790–1,050
Curacoa 6	ANU 8487, 8488	1,100 $\pm$ 50	930–1,150
Curacoa 7	ANU 8490–8492, 9991	1,475 $\pm$ 32	1,310–1,480
Curacoa 8	ANU 8493–8495, 8497	1,675 $\pm$ 45	1,480–1,700
Curacoa 9	ANU 8505, 9984, 9987	2,065 $\pm$ 40	1,930–2,150
Curacoa 10	ANU 8519–8521, 8498	2,240 $\pm$ 30	2,150–2,340
Curacoa 11	ANU 9989, 9992	2,540 $\pm$ 50	2,410–2,760
Curacoa 12	ANU 8504, 8517, 8518	2,830 $\pm$ 40	2,850–3,070
Curacoa 13	ANU 8501, 9994	3,025 $\pm$ 55	3,050–3,360
Curacoa 14	ANU 8502, 8506, 8507	3,360 $\pm$ 40	3,480–3,710
Curacoa 15	ANU 8516, 9997	3,640 $\pm$ 50	3,840–4,130
Curacoa 16	ANU 8514, 9996	3,780 $\pm$ 55	3,990–4,400
Curacoa 17	ANU 8547–8549	4,070 $\pm$ 75	4,410–4,820
Curacoa 18	ANU 8508, 9998	4,230 $\pm$ 55	4,580–4,960
Curacoa 19	ANU 8510, 9999	4,410 $\pm$ 55	4,870–5,270
Curacoa 20	ANU 8511, 10000	4,820 $\pm$ 55	5,340–5,720
Curacoa 21	ANU 8512	5,095 $\pm$ 80	5,660–6,090
Curacoa 22	ANU 8513, 10001	5,620 $\pm$ 60	6,310–6,610

Note progressive increase in age for each ridge/storm deposit inland. Ridge 1 at each location is the seaward ridge. All values in the radiocarbon age column are group mean reservoir-corrected ages, except for Wk ages which are group mean conventional ages.

been less intense (929–931 hPa) than elsewhere; however, the lower elevation of these ridges may be due to limited sediment supply²² rather than less intense cyclones.

The frequency of category 5 cyclones in the GBR region has previously been assumed to be once every several millennia^{7,12–14}.

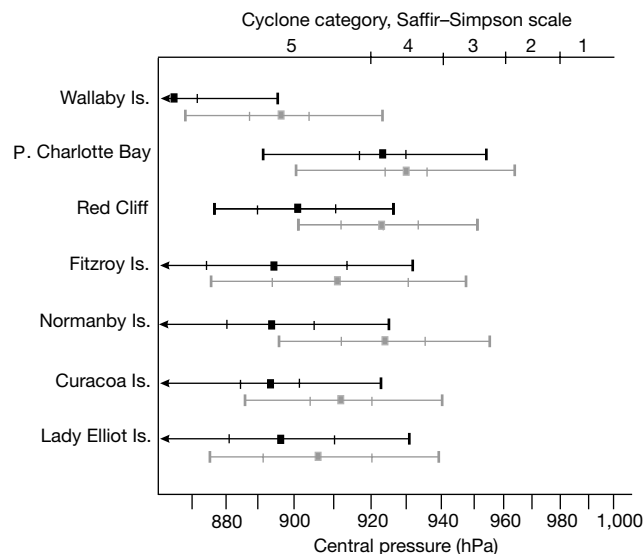


Figure 2 Central pressures (mean, 1 σ , 2 σ) of cyclones responsible for building ridges and eroding terraces at study sites. Black lines represent cyclone intensities responsible for the emplacement of most ridges at sites. Grey lines indicate less-intense cyclones responsible for constructing and eroding lower-elevation ridges and terraces, respectively. Central pressures reported here must be regarded as minimum intensity values, as ridge heights represent the minimum elevation of storm inundation (see Methods).

Our results suggest that these events occur every 200–300 yr within all regions of the GBR between latitudes 13° and 24° S, and also within the Gulf of Carpentaria. We believe that our sites are representative of the long-term frequency of extreme intensity cyclones as they occur across a wide variety of environments along virtually the entire GBR. Our modelling results also agree closely with the levels of inundation and central pressures recorded during historical cyclones at Red Cliff (category 2 events in 1997 and 2000), Curacoa (category 3 event in 1971) and near PCB (category 5 event in 1899).

Extreme-intensity cyclones can cause destruction of substantial areas of rainforest and coral reefs. Large-scale severe wind disturbances elsewhere have resulted in forest gaps thousands of hectares in area, with up to 87–100% tree mortality with subsequent recruitment and succession^{18,23}. More severe storms, such as the super-cyclones occurring every two to three centuries in northeast

Table 2 Central pressure of tropical cyclones responsible for emplacing ridges and eroding terraces

Location	Inun. (m)	H_s (m)	Set-up (m)	Run-up (m)	Surge (m)	Mean central pressure (hPa)	$\pm 1\sigma$	$\pm 2\sigma$
Wallaby	5.08	3.9	0.39	1.2	3.43	861	9	30
Wallaby	4.1	3.6	0.36	1.1	2.58	896	8	29
Wallaby	3.6	3.4	0.38	1.0	2.24	910	8	29
PCB	3.1	2.32	0.23	0.7	2.17	924	9	30
PCB	2.9	2.26	0.22	0.67	1.99	931	9	29
Red Cliff	6.1	5.2	0.52	1.56	3.99	900	12	25
Red Cliff	4.9	4.7	0.47	1.4	3.0	926	12	24
Red Cliff	3.8	4.13	0.41	1.24	2.11	949	12	23
Red Cliff	2.12	3.1	0.31	0.92	0.86	982	10	17
Fitzroy	4.5	7.4	0.74	2.2	1.5	894	20	39
Fitzroy	3.9	6.7	0.67	2.0	1.2	912	19	36
Normanby	4.71	7.9	0.79	2.4	1.4	893	13	31
Normanby	3.8	6.7	0.67	2.0	1.0	924	12	29
Curacoa	5.5	6.3	0.63	1.89	2.95	893	9	29
Curacoa	4.8	5.8	0.58	1.73	2.47	912	8	28
Lady	5.1	11.2	1.12	3.4	0.52	896	15	35
Lady	4.9	10.6	1.06	3.2	0.47	906	14	32
Lady	4.5	9.9	0.99	3.0	0.43	916	14	30

'Lady', Lady Elliot Island; Inun., inundation or height of ridge or terrace; H_s , significant wave height. Lower inundation heights represent heights of lower storm deposits/ridges and terraces. Mean central pressure represents cyclone intensity if storm crossed at mean tide level, which is the dominant tide level over the full nodal (~19-yr) tidal cycle. Central pressures are regarded as minimum intensity values (see Methods).

Australia, would likewise result in large areas of tree mortality, uprooting and eventual succession.

The damage incurred by coral communities from severe intensity cyclones can also be devastating⁵. There is substantial spatial variation in damage following cyclones^{1,5,24,25}, and storms of different intensity can have varied effects⁵. However, even in rare cases where extreme cyclones do not cause extensive damage to coral communities, they weaken the substrate, allowing severe damage to occur during subsequent less intense events²⁵. Hence the frequency of the most intense cyclones is critical to the level of damage experienced by these communities over the longer term. The frequency of these extreme events may also place constraints upon the longevity of individual corals. Whereas some corals live up to 700 yr, most corals on the GBR rarely live beyond a few centuries²⁶. Only the highest storm waves are capable of dislodging the largest of these long-lived corals, and such extreme conditions are likely to occur only during super-cyclones. That these events occur at time intervals less than the life span of many individual corals and trees suggests that the character of these communities must be strongly influenced by these high-magnitude storms.

The inferred frequency of super-cyclones is sufficiently high to have serious implications for the vulnerability of coastal human settlements. The twentieth century in northeast Australia was devoid of any category 5 cyclones, with only one such event (1899) since European settlement in the mid nineteenth century. The GBR region, however, experienced at least five such storms over the past 200 yr, with the area now occupied by Cairns experiencing two super-cyclones between 1800 and 1870 (Table 1). The periodicity of these storms suggests that community vulnerability, exposure and risk is much higher than previously estimated.

Short-term historical records in this region, and probably elsewhere throughout the global tropics, are inadequate for determining future effects of tropical cyclones on human society and their role as disturbance mechanisms upon ecological communities. Assessing these effects requires decoupling the role of human-induced climate change from natural quasi-periodicities, and this can only be achieved through the examination of longer-term records. □

Methods

Surge heights were modelled using GCO2MD²⁷—a numerical model that solves a set of mathematical equations over an equally spaced grid to determine water depth, currents, topography and bathymetry, and incorporates wind stresses and atmospheric pressure gradients acting on the ocean surface, and friction on the ocean floor. These results were compared with the maximum envelope of oceanic waters (MEOW) model²⁸ and with an empirical model—the Jelesnianski–Trajer nomogram technique²⁹: these three models produced similar results. The relationship between surge height and angle of cyclone approach, cyclone translational velocity, radius of maximum winds (R_m) and central pressure was determined for each site. Surge was found to increase with increases in each of these factors, but because their respective magnitudes are unknown for the palaeo-event, their mean probability magnitude level was calculated from the historical record of tropical cyclones in the region. The mean translational velocity of cyclones in this region is $15.5 \pm 6.6 \text{ km h}^{-1}$, R_m is $30 \pm 10 \text{ km}$, and cyclone approach angle is $63.8^\circ \pm 27.1^\circ$ (ref. 14). We used the 99% probability level of occurrence for the velocity and R_m (translational velocity 40 km h^{-1} , $R_m = 60 \text{ km}$), and an approach angle of 37° (being the angle producing the highest possible surge) for all of our surge model analyses. All modelled cyclones crossed the coast with the zone of greatest wind velocity occurring generally less than 5 km from each of the shingle ridge or terrace study sites.

Wave characteristics were determined using a nearshore shallow water wave model incorporating the GBR and local bathymetry and run in conjunction with the surge model. Wave set-up was estimated at 10% of significant wave height (H_s), and run-up determined at 30% of H_s , on the basis of observations of run-up during three cyclone events in March 1997, February 1999 and February 2000 at Red Cliff and Fitzroy island. These observations involved the recording of total inundation by pre-placed markers at the sites, both the modelled and direct tide gauge measurements of the surge height, calculation of the wave set-up at 10% of H_s and direct measurement of H_s by an offshore wave-rider buoy in water of depth 15–20 m. The mean central pressure of the cyclone was determined as follows: the central cyclone pressure that produces 30% of H_s (measured run-up) and 10% of H_s (calculated wave set-up) is equal to the central cyclone pressure that produces the necessary surge height; the sum of these three components of the water column (run-up, set-up and surge) plus the tide level is equivalent to the height of the terrace or shingle ridge. Uncertainty margins reflect possible 1σ and 2σ tide range, determined from frequency distribution nodal tide curve for each site.

Cyclone frequency for PCB, Curacao, Lady Elliot and Wallaby islands was calculated from the average interval in years between ridge-building or storm-deposit events from the youngest to the oldest ridges dated from ^{14}C determinations. A total of 40 dates from 9 ridges at Lady Elliot island, 68 dates from 16 ridges at Curacao island and 65 dates from 12 ridges at PCB show no change in frequency in ridge-building events over the past several millennia, and a progressive increase in age of each ridge inland^{8–11} (Fig. 1b). Ascertaining the most recent inundations of the ridges at Fitzroy and Normanby islands required collection of coral fragments from ridge crests. Although both sites returned recent ages (after AD 1800, it is highly unlikely that ridges at each site were deposited during the same event as coral fragments from the landward ridge were covered in lichen (shading by trees and so on has not influenced lichen growth) and had a more weathered appearance. The landward ridge is therefore interpreted as an older (and separate) storm deposit than its seaward counterpart. Grain size analyses of the gravels comprising the terraces at Red Cliff show they are well sorted, and (apart from the lower terrace) show no sign of grain size variations from the toe to the crest of each terrace. This indicates that the upper three terraces are erosional features, and that the deposit is a single depositional unit. As coral fragments younger than AD 1800 are buried in the top terrace, it can be assumed each of the lower terraces are younger again. Ridge heights represent the minimum level of inundation during a cyclone event, and therefore our estimates of palaeo-cyclone intensity must be regarded as minima. Indeed, morphological and sedimentological evidence at the multiple ridge sites suggest that these ridge sequences have been compacted over time and regularly over-topped by storm inundation since deposition, effectively reducing the original ridge height.

Received 20 February; accepted 21 August 2001.

- Connell, J. Diversity in tropical rainforests and coral reefs. *Science* **199**, 1302–1310 (1978).
- Sousa, W. The role of disturbance in natural communities. *Annu. Rev. Ecol. Syst.* **15**, 353–391 (1984).
- Tanner, J. *et al.* The role of history in community dynamics: A modelling approach. *Ecology* **77**, 108–117 (1996).
- Hughes, T. & Connell, J. Multiple stressors on coral reefs: A long-term perspective. *Limnol. Oceanogr.* **44**, 932–940 (1999).
- Connell, J. *et al.* A 30 year study of coral abundance, recruitment and disturbance at several scales in space and time. *Ecol. Monogr.* **67**, 461–488 (1997).
- Oliver, J. & Britton, N. *Natural Hazards and Reinsurance* (Cumberland College Health Sciences, Sydney, Australia, 1989).
- Walsh, K. *et al.* *Climate Change in Queensland under Enhanced Greenhouse Conditions* (Third annual report 1999–2000, CSIRO Atmospheric Research, Aspendale, Victoria, Australia, 2001).
- Chappell, J. *et al.* Holocene palaeo-environmental changes, central to north Great Barrier Reef inner zone. *BMR J. Aust. Geol. Geophys.* **8**, 223–235 (1983).
- Chivas, A. *et al.* Radiocarbon evidence for the timing and rate of island development, beach rock formation and phosphatization at Lady Elliot Island, Queensland, Australia. *Mar. Geol.* **69**, 273–287 (1986).
- Hayne, M. & Chappell, J. Cyclone frequency during the last 5,000 yrs from Curacao Island, Queensland. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **168**, 201–219 (2001).
- Hayne, M. & Chappell, J. A record of cyclone frequency from Princess Charlotte Bay, Nth Qld, Australia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* (in the press).
- Lourens, H. Tropical cyclones in the Australian region, July 1909 to June 1980. (Bureau of Meteorology, Melbourne, 1981).
- Harper, B. *Storm Tide Threat in Queensland: History, Prediction and Relative Risks* (Technical report 10, Queensland Department of Environment & Heritage, Brisbane, 1998).
- McInnes, K. *et al.* *Cyclones Impacts on Tourism, Cairns, Nth Queensland* (CSIRO Report, CSIRO Tourism, Melbourne, 2000).
- Hubbell, S. *et al.* Light gap disturbances, recruitment limitation, and tree diversity in a neotropical rainforest. *Science* **283**, 554–557 (1999).
- Bryant, E. *Natural Hazards* (Cambridge Univ. Press, Cambridge, 1991).
- Henderson-Sellers, A. *et al.* Tropical cyclones and global climate change; a post IPCC assessment. *Bull. Am. Meteorol. Soc.* **79**, 19–38 (1998).
- Blumenstock, D. Typhoon effects at Jaluit Atoll in the Marshall Islands. *Nature* **182**, 1267–1269 (1958).
- Maragos, J. *et al.* Tropical cyclone creates a new land formation on Funafuti Atoll. *Science* **181**, 1161–1164 (1973).
- Fitchett, K. Physical effects of Hurricane Bebe upon Funafute Atoll, Tuvalu. *Aust. Geographer* **18**, 1–7 (1987).
- Scoffin, T. The geological effects of hurricanes on coral reefs and the interpretation of storm deposits. *Coral Reefs* **12**, 203–221 (1993).
- Chappell, J. & Grindrod, J. in *Coastal Geomorphology in Australia* (ed. Thom, B.) 197–231 (Academic, Sydney, 1984).
- Everham, E. & Brokaw, N. Forest damage and recovery from catastrophic wind. *Bot. Rev.* **62**, 113–184 (1996).
- Hughes, T. Community structure and diversity of coral reefs: The role of history. *Ecology* **7**, 275–279 (1989).
- Lirman, D. & Fong, P. Susceptibility of coral communities to storm intensity, duration and frequency. *Proc. 8th Int. Coral Reef Symp.* (eds Lessios, H. & MacIntire, I.) Vol. 1, 561–566 (Smithsonian Tropical Research Institute Balboa, Panama, 1997).
- Potts, D. C. *et al.* Dominance of a coral community by the genus *Porites*. *Mar. Ecol. Prog. Ser.* **23**, 79–84 (1985).
- Hubbert, G. & McInnes, K. A storm surge inundation model for coastal planning and impact studies. *J. Coast. Res.* **15**, 168–185 (1999).
- Sanderson, B. *et al.* *A Tropical Cyclone Maximum Envelope of Waters (MEOW) Technique* (Applied Mathematics Report 95/28, Department of Mathematics, Monash Univ. Clayton, Victoria, Australia, 1995).
- US Army Corps of Engineers *Shore Protection Manual* (US Army Coastal Engineering Research Centre, Washington, 1986).

Acknowledgements

We thank J. Wright, J. Tanner and H. Marsh for comments on the manuscript; T. Hughes, J. Chappell, G. Hubbert and S. Oliver for discussions on the topic; and J. H. Choat and R. Pearson for early encouragement of this project. M.H. publishes with the permission of the Chief Executive Officer of AGSO.

Correspondence and requests for material should be addressed to J.N.
(e-mail: Jonathan.Nott@jcu.edu.au).

Multiple benefits of gregariousness cover detectability costs in aposematic aggregations

Marianna Riipi*, Rauno V. Alatalo, Leena Lindström & Johanna Mappes

University of Jyväskylä, Department of Biological and Environmental Science, Konnevesi Research Station, PO Box 35, FIN-40 351 Jyväskylä, Finland

Understanding the early evolution of aposematic (warning) coloration has been a challenge for scientists, as a new conspicuous morph in a population of cryptic insects would have a high predation risk and would probably die out before local predators learnt to avoid it^{1–4}. Fisher⁵ presented the idea of aggregation benefit through the survival of related individuals; however, his theory has been strongly debated^{6–8} as the mechanisms that favour grouping have never been explored experimentally with the incorporation of detectability costs. Here we create a comprehensive ‘novel world’ experiment with the great tit (*Parus major*) as a predator to explore simultaneously the predation-related benefits and costs for aposematic aggregated prey, manipulating both group size and signal strength. Our results show that grouping would have been highly beneficial for the first aposematic prey individuals surrounded by naive predators, because (1) detectability risk increased only asymptotically with group size; (2) additional detectability costs due to conspicuous signals were marginal in groups; (3) even naive predators deserted the group after detecting unpalatability (dilution effect); and (4) avoidance learning of signal was faster in groups. None of these mechanisms require kin selection.

In the first experiment we tested how signal strength and group size affects the risk of prey detection. We used the ‘novel world’ method^{6,9} because it uses a fundamental property of warning signals, conspicuousness, but presents the predators with signals not found in their natural environment. As prey, we used pieces of almond glued between two pieces of paper that had the signal printed on them. The birds were simultaneously presented with nine different palatable prey assemblages—combinations of three group sizes (1, 4 or 8 prey items) and three signals (a cross symbol with a square set at the centre: the size of the square varied according to the signal number, that is, small (signal 2) to large (signal 4); see Fig. 1). The background on the aviary floor consisted of white paper sheets with the cryptic signal (signal 1: a cross symbol without a square) printed on them. Each prey assemblage was presented four times in random locations on the background. Increasing group size caused an increase in the number of those assemblages that were attacked by the birds ($F = 84.32$, degrees of freedom (d.f.) = 2; $P < 0.001$), owing to an increase in prey detectability. Prey groups had higher detectability risk than solitary prey, but this risk did not increase in direct proportion to group size (this effect is sometimes called concealment¹⁰ or avoidance effect¹¹) (Fig. 2).

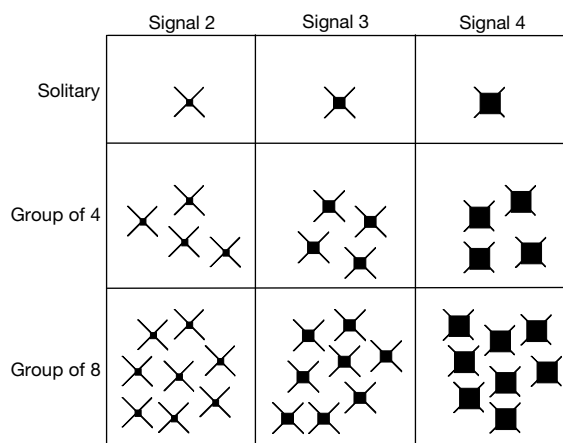


Figure 1 Prey assemblages used in the detectability experiment. Each assemblage was presented in the aviary in four replicates, and thus there were 36 assemblages in total.

Increasing signal strength caused an increase in the number of those assemblages that were attacked by the birds ($F = 9.91$, d.f. = 2, $P < 0.001$). For solitary prey, a strong signal (signal 4) increased the detectability of the prey twofold (100%) compared with signal 2 (Fig. 2; t -test between solitary signal 2 and 4, $t = 3.80$, d.f. = 64, $P = 0.001$ after Bonferroni correction), whereas in groups of four or eight prey items the increase in the detectability of the group caused by the strong signal was only 12.5% and 13.3%, respectively (Fig. 2; group of 4: $t = 1.34$, d.f. = 64, $P = 0.561$; group of 8: $t = 1.69$, d.f. = 64, $P = 0.288$ after Bonferroni correction). Thus, signal strength caused little additional increase in detectability when in a group, suggesting that the cost of a strong signal could be smaller for aggregations than for solitary prey. Previously, the effect of the strength of an aposematic signal on the detectability of prey has been assumed to be similar for both a small group (under 20 individuals) and for solitary prey^{10,12}.

If the increased detectability of a group compared with solitary prey does not translate into increased mortality risk per individual, then grouping is always more favourable than a solitary lifestyle. A decreased per capita mortality risk can be produced by a dilution effect, which means that when a predator cannot eat the whole group, an individual's chance of being eaten is smaller in a group than when it is solitary^{10,11,13}. To show the effect of dilution, we used the detectability data (Fig. 2) to calculate estimates of relative mortality for average prey individuals in different group sizes in situations where the predator tastes a certain number of prey items from a group (Table 1). This simulates different levels of predator satiation¹², for example, if the prey is very unpalatable the predator may leave the group after tasting only one individual. Naturally, the real group sizes favoured by selection are dependent on the predator and prey in question, but the general idea is that prey individuals with strong signals benefit most from gregariousness, because their detectability is increased only slightly by grouping (Table 1).

In the second experiment, we performed learning tests with six groups of birds in a factorial design with two signal levels (signal 2 (weak) and signal 4 (strong)) and three prey group sizes. Each bird was presented with solitary, palatable cryptic items (the symbol on the prey was similar to the background symbol) together with unpalatable prey items carrying a signal of only one type. Solitary cryptic prey were palatable to mimic a situation where a distasteful aposematic morph evolves in an environment where most of the cryptic species are palatable; there is thus a reward of learning to avoid the aposematic prey. The tests were carried out in five consecutive days for each bird, and numbers of cryptic and aposematic prey eaten at the trials were recorded. Survival of unpalatable prey with the weak signal (signal 2) increased with group size (Fig. 3a; $F_{(2,20)} = 3.74$, $P = 0.042$). The birds did not learn to avoid

* Present address: Section of Ecology, Department of Biology, University of Turku, FIN-20014 Turku, Finland.

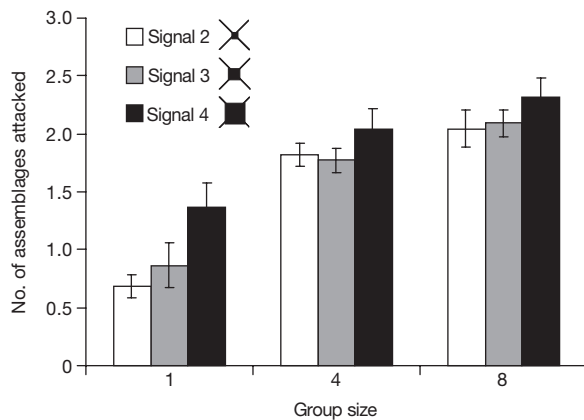


Figure 2 Detectability risk of each prey assemblage as the number of attacked prey assemblages in the detectability experiment. Error bars indicate s.e.m.

unpalatable prey, as trial had no effect on survival ($F_{(4,17)} = 2.10$, $P = 0.126$). There was no interaction between learning (trial) and group size ($F_{(8,36)} = 0.44$, $P = 0.890$). Thus, the birds did not learn to use either the signal (no learning in solitary treatment) or gregariousness alone (no learning in group treatments) as a sign of unpalatability. As with the weaker signal, survival of the prey with signal 4 increased with group size (Fig. 3b; $F_{(2,22)} = 9.10$, $P = 0.001$). However, the birds learnt to avoid the aposematic prey, which increased the survival of the aposematic prey in the later trials ($F_{(4,19)} = 17.95$, $P < 0.001$). Learning was faster with bigger groups of prey items (Fig. 3b; interaction between learning and group size: $F_{(8,40)} = 2.18$, $P = 0.050$).

We combined the two experiments to compare the observed mortality of unpalatable prey in the learning experiment with the mortality that would be expected if the birds ate the aposematic prey according to their detectability. From the detectability risks of different prey assemblages (Fig. 2), we calculated expected mortality values (horizontal lines in Fig. 3) for the learning experiment, assuming that all the prey items would be eaten from a group once it was found. The birds ate distasteful prey items that displayed signal 2 according to their detectability only when the prey were solitary (Fig. 3a). When the prey items were gregarious, they suffered lower mortality than expected on the basis of their detectability. The difference between the expected and observed mortality of prey increased with increasing group size (Fig. 3a). The enhanced survival of aggregated unpalatable prey with a weak signal was probably due to a dilution effect, as a result of the birds leaving the group after eating one or a few prey items and detecting unpalatability. Dilution is beneficial only if the encounter probability of a group is smaller than that of an equal number of solitary prey

Table 1 Effect of dilution on per capita mortality risks for prey individuals in different group sizes

Signal	Group size	Average mortality risk for an individual				
		1 eaten	2 eaten	3 eaten	4 eaten	8 eaten
Signal 2	4	0.67	1.33	2	2.67	—
	8	0.38	0.75	1.13	1.50	3
Signal 3	4	0.51	1.03	1.54	2.05	—
	8	0.30	0.61	0.91	1.21	2.42
Signal 4	4	0.38	0.75	1.13	1.50	—
	8	0.21	0.43	0.64	0.85	1.70

Risks were calculated relative to the detectability of the similar solitary prey (defined as 1) within each signal. Thus, comparisons should be made only between group sizes within a signal, as the values for different signals are not comparable. The original values for the groups are the attack numbers from the detectability experiment (relative to the detectability of the solitary item within each signal), which are divided by the number of prey in the group to get a theoretical risk for one individual and multiplied by the number of prey eaten in each case. For example, even if four items are eaten from a group of eight with signal 4, it is better for the individuals with this signal to live in a group of eight than solitary, because the mortality risk per average individual is lower in the group (0.85) than solitary (1.0). Numbers less than 1 indicate when it is advantageous for prey to live in a group rather than solitary.

items¹¹. This was the case here because in the detectability experiment the detectability risk of a group did not increase linearly with group size. Because the dilution effect is based on proximity cues and works before the predator has learnt to avoid a group on the basis of the signal, it can also help cryptic unpalatable prey that live gregariously. Signal 2 in these experiments was effectively cryptic, as it did not differ from signal 1 in a visibility test with solitary prey⁹, and it did not cause avoidance learning in the birds even after repeated experience.

Aposematic prey items with a strong signal (signal 4) were eaten less than predicted from their detectability in all group sizes, even in the first trial (Fig. 3b). This might be due to a neophobic reaction towards unfamiliar prey¹⁴, because to birds the prey item with a strong signal might be visually more different from the white training baits than the cryptic prey item. The expected mortality was based on the detectability test, in which the birds had previous experience with all the signal types and did not show any initial avoidance. In aggregations, the difference between expected and observed mortality in the first trial was too large to be explained only by neophobia (compare the first trial in solitary treatment with the first trial in group treatments, Fig. 3b). This suggests that dilution was also functioning with the stronger signal.

We demonstrate that gregariousness has more benefits for aposematic prey than just faster avoidance learning^{6,15} and possible kin benefits⁵. For the more conspicuous signal, gregariousness provided many advantages that lead to the better survival of the signal

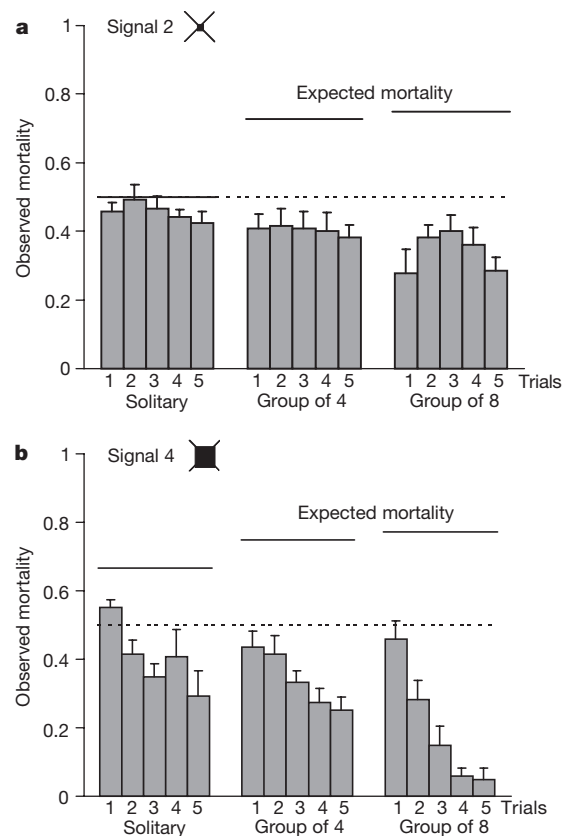


Figure 3 Observed and expected mortality of unpalatable prey as the proportion of unpalatable items from all prey that were eaten. Bars represent observed mortality (with standard error) for each treatment in the five trials. The dashed line represents equal mortality for palatable cryptic and unpalatable aposematic prey. The three solid horizontal lines indicate expected mortality for each group size estimated from the detectability risks in Fig. 2. Expected mortality was calculated relative to the detectability of solitary prey displaying signal 2, assuming that all prey items in a group were eaten once it was found. **a**, Signal 2. **b**, Signal 4.

carriers. Avoidance learning, which was faster with the larger aggregation, required a strong signal, but the signal had no initial cost when the prey were gregarious. The signal increased detectability only slightly and detectability costs resulting from both the signal and aggregation were counterbalanced by the dilution effect, by decreased per capita encounter probability and possibly by neophobia. Notably, the dilution effect increased the survival of aggregated unpalatable prey even without a strong warning signal. Thus, unpalatability alone could select for grouping under the influence of individual selection, and in groups the evolution of a stronger signal would be favoured by synergistic selection^{6,16}, which affects individuals of the same phenotype regardless of their ancestral relatedness. Alternative explanations for the initial evolution of warning coloration have been proposed¹⁷, such as random drift, neophobia¹⁴, evolution through individual selection when prey are by some means able to survive an attack^{18–20}, or the coloration being cryptic from a distance but aposematic when the predator is close²¹. Most of these hypotheses are not mutually exclusive, and different mechanisms may have been important in distinct areas or with different species, thus we do not claim that gregariousness is a prerequisite for the evolution of warning signals. However, given the dilution effect, the small detectability costs of signals in groups and the enhanced learning of strong signals in groups, it seems that gregariousness of unpalatable prey might have enabled the initial appearance of aposematism, and grouping may assist in the survival of established aposematic prey whenever the prey encounter naive predators. □

Methods

Predators and prey

Wild great tits were caught in mist nets around Konnevesi Research Station where the experiments were carried out from January to May 1997 (general methods as in ref. 9). Each bird was trained to open similar paper prey items to the ones that were eventually used in the experiments, but during the training the prey items had no signal. After the experiment we released great tits to the area where they were caught.

Detectability experiment

All of the prey items were palatable as the objective was to find out how group size and signal conspicuousness affect the number of prey attacked owing to detectability differences. Signal 1 (the background signal) was not used because in a separate visibility test with solitary prey items, signal 1 and signal 2 did not differ significantly in their conspicuousness to the great tits (result reported in ref. 9). Before the detectability test, the birds ($n = 11$) were given palatable prey items that displayed all of the signals used so as to avoid neophobic reactions towards any of the signals. Eating or touching the prey item was taken as an indication that the bird had seen the prey, as the birds had no reason to avoid any of the prey types. The trial continued until the bird had attacked 20 prey assemblages, but only the first 15 assemblages were included in the final analysis to avoid the risk that prey depletion during feeding would bias the detectability estimation. The experiment was repeated the next day and the mean values from two trials were used in further calculations, because the trials gave similar results. The data were analysed with a two-way analysis of variance (ANOVA) with main effects (SPSS for Windows version 7.0).

Learning experiment

We tested selection pressures on evolving aposematic prey by using palatable cryptic prey items with signal 1 together with unpalatable items with either signal 2 or signal 4. Signal 3 was not included because it did not differ much from signal 2 (see Fig. 2). Each bird was randomly assigned to one of the six treatment groups (two signal strengths $\times$ three group sizes), so that every treatment had 7–9 birds (total $n = 48$ birds).

In each treatment half of the prey items in the aviary were cryptic, palatable and solitary, whereas the other half were aposematic (unpalatable and displaying a signal) and were placed either solitarily, in groups of four or in groups of eight. The number of prey items was always the same: 24 palatable and 24 distasteful items in the aviary. The birds were allowed to taste 15 prey items in each of the five trials. The number of unpalatable prey items eaten in a trial was used as a dependent variable in repeated measures ANOVA (SPSS for Windows version 7.0), with trial as the within-subject factor (corresponding to learning) and group size as a between-subject factor. Separate tests were performed for the two signals. The data for solitary treatment, which serves as the control here, were obtained from another experiment that was performed at the same time⁹.

Received 9 February; accepted 13 August 2001.

1. Endler, J. A. Frequency-dependent predation, crypsis and aposematic coloration. *Phil. Trans. R. Soc. Lond. B* **319**, 505–523 (1988).
2. Guilford, T. in *Insect Defenses. Adaptive Mechanisms and Strategies of Prey and Predators* (eds Evans, D. L. & Schmidt, J. O.) 23–61 (State Univ. New York Press, New York, 1990).

3. Mallet, J. & Singer, M. C. Individual selection, kin selection, and the shifting balance in the evolution of warning colours: the evidence from butterflies. *Biol. J. Linn. Soc.* **32**, 337–350 (1987).
4. Schuler, W. & Roper, T. J. Responses to warning coloration in avian predators. *Adv. Study Behav.* **21**, 111–146 (1992).
5. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
6. Alatalo, R. V. & Mappes, J. Tracking the evolution of warning signals. *Nature* **382**, 708–710 (1996).
7. Tullberg, B. S., Leimar, O. & Gamberale-Stille, G. Did aggregation favour the initial evolution of warning coloration? A novel world revisited. *Anim. Behav.* **59**, 281–287 (2000).
8. Alatalo, R. V. & Mappes, J. Initial evolution of warning coloration: comments on the novel world method. *Anim. Behav.* **60**, F1–F2 (2000).
9. Lindström, L., Alatalo, R. V., Mappes, J., Riipi, M. & Vertainen, L. Can aposematic signals evolve by gradual change? *Nature* **397**, 249–251 (1999).
10. Treisman, M. Predation and the evolution of gregariousness. I. Models for concealment and evasion. *Anim. Behav.* **23**, 779–800 (1975).
11. Turner, G. F. & Pitcher, T. J. Attack abatement: a model for group protection by combined avoidance and dilution. *Am. Nat.* **128**, 228–240 (1986).
12. Sillén-Tullberg, B. & Leimar, O. The evolution of gregariousness in distasteful insects as a defense against predators. *Am. Nat.* **132**, 723–734 (1988).
13. Hamilton, W. D. Geometry for the selfish herd. *J. Theor. Biol.* **31**, 295–311 (1971).
14. Lindström, L., Alatalo, R. V., Lyytinen, A. & Mappes, J. Predator experience on cryptic prey affects the survival of conspicuous aposematic prey. *Proc. R. Soc. Lond. B* **268**, 357–361.
15. Gagliardo, A. & Guilford, T. Why do warningly-coloured prey live gregariously? *Proc. R. Soc. Lond. B* **251**, 69–74 (1993).
16. Guilford, T. Is kin selection involved in the evolution of warning coloration? *Oikos* **45**, 31–36 (1985).
17. Mallet, J. & Joron, M. Evolution of diversity in warning color and mimicry: polymorphisms, shifting balance, and speciation. *Annu. Rev. Ecol. Syst.* **30**, 201–233 (2000).
18. Järvi, T., Sillén-Tullberg, B. & Wiklund, C. The cost of being aposematic. An experimental study of predation on larvae of *Papilio machaon* by the great tit *Parus major*. *Oikos* **36**, 267–272 (1981).
19. Sillén-Tullberg, B., Wiklund, C. & Järvi, T. Aposematic coloration in adults and larvae of *Lygaeus equestris* and its bearing on mullerian mimicry: an experimental study on predation on living bugs by the great tit *Parus major*. *Oikos* **39**, 131–136 (1982).
20. Marples, N. M., van Veelen, W. & Brakefield, P. M. The relative importance of colour, taste and smell in the protection of an aposematic insect *Coccinella septempunctata*. *Anim. Behav.* **48**, 967–974 (1994).
21. Weismann, A. *Studies in the Theory of Descent* (Sampson, Low, Marston, Searle and Rivington, London, 1882).

Acknowledgements

We thank H. Nisu, L. Vertainen and the Academic Hobby Crafts Club for their help, and Konnevesi Research Station for the facilities. We also thank A. Chaine, G. Corrigan, A. Kause, R. Leimu, B. Lyon, A. Lyytinen, J. Tuomi and B. Weaver for valuable comments on earlier versions of the manuscript. This study was supported by the Academy of Finland. Authors after M.R. are in alphabetical order.

Correspondence and requests for materials should be addressed to M.R. (e-mail: marianna.riipi@utu.fi) or J.M. (e-mail: mappes@cc.jyu.fi).

Positive selection of a gene family during the emergence of humans and African apes

Matthew E. Johnson*, Luigi Viggiano†, Jeffrey A. Bailey*, Munah Abdul-Rauf‡, Graham Goodwin‡, Mariano Rocchi† & Evan E. Eichler*

* Department of Genetics and Center for Human Genetics, Case Western Reserve University School of Medicine and University Hospitals of Cleveland, Cleveland, Ohio 44106, USA

† DAPEG, Sezione di Genetica, Via Amendola 165/A, 70126 Bari, Italy

‡ Section of Molecular Carcinogenesis, Institute of Cancer Research, Haddow Laboratories, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK

Gene duplication followed by adaptive evolution is one of the primary forces for the emergence of new gene function¹. Here we describe the recent proliferation, transposition and selection of a 20-kilobase (kb) duplicated segment throughout 15 Mb of the short arm of human chromosome 16. The dispersal of this segment was accompanied by considerable variation in chromosomal-map location and copy number among hominoid species. In humans, we identified a gene family (*morpheus*) within the duplicated segment. Comparison of putative protein-encoding

exons revealed the most extreme case of positive selection among hominoids. The major episode of enhanced amino-acid replacement occurred after the separation of human and great-ape lineages from the orangutan. Positive selection continued to alter amino-acid composition after the divergence of human and chimpanzee lineages. The rapidity and bias for amino-acid-altering nucleotide changes suggest adaptive evolution of the *morpheus* gene family during the emergence of humans and African apes. Moreover, some genes emerge and evolve very rapidly, generating copies that bear little similarity to their ancestral precursors. Consequently, a small fraction of human genes may not possess discernible orthologues within the genomes of model organisms.

During physical mapping and sequencing of the human genome²⁻⁵, a complex series of duplicated genomic segments were identified that mapped to multiple cytogenetic band positions on chromosome 16 (Fig. 1a). We reassessed the genomic distribution and the extent of duplication of one of these segmental duplications, termed LCR16a (low-copy repeat sequence 'a' from chromosome 16)³. Fifteen distinct copies of the duplicated segment were characterized (see Methods and Supplementary Information Fig. 1). These genomic repeats were specific to human chromosome 16, were ~20 kb long, and shared a remarkably high degree of sequence identity (Fig. 1b). Searches for sequence similarity in the expressed sequence divisions of GenBank revealed a previously uncharacterized family of genes within the LCR16a segment (Fig. 1c and see Supplementary Information). Transcripts could be identified for 6

of the 15 genomic copies. We found no significant sequence similarity to this gene family in other organisms either at the nucleotide or protein level (sequence similarity values $E < 10^{-30}$ and $E < 10^{-2}$, respectively), indicating a highly diverged family of human transcripts. Sequence comparison of putative proteins from two full-length human transcripts showed 81% amino-acid sequence identity (see Supplementary Information Fig. 2). In sharp contrast, the corresponding non-coding portions of genomic DNA were 98.1% identical. These data suggested either that the exonic regions were hypermutable or that amino-acid changes had been selected during the evolution of this gene family.

The high degree of genomic sequence similarity among the various human copies (~98%) indicated a recent evolutionary divergence. Analysis by fluorescence *in situ* hybridization (FISH) of primate metaphase chromosomes and interphase nuclei confirmed marked variation in signal intensity, copy number and map location (Fig. 2). Among all Old World monkeys, a single metaphase signal corresponding to one or two copies (by interphase nuclei) was identified distally on chromosome 16. Sequence analysis of a genomic subclone from baboons (data not shown) confirmed an orthologous map position to human sequence 16p13.1—the probable ancestral segment from which all other copies originated. In contrast to the Old World monkeys, the genome of the great apes showed a major proliferation of the LCR16a duplication. This is particularly evident within the short arm, which is almost completely 'painted' by the LCR16a probe. This effect is most striking in the lineages of humans and African great apes. Using a combination

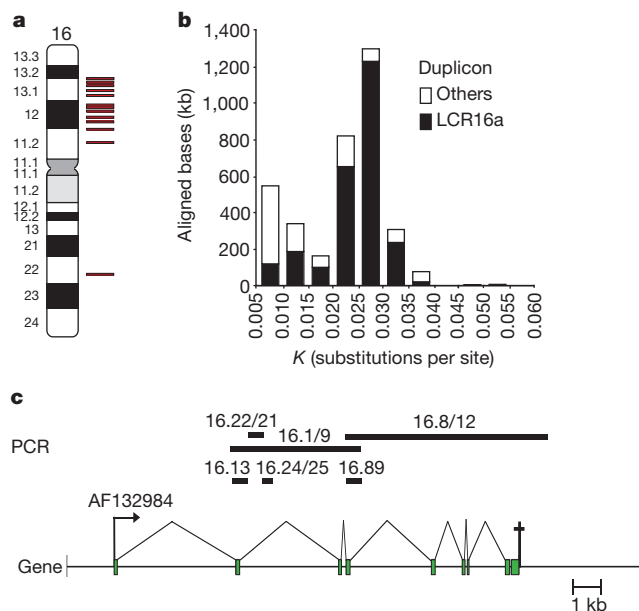


Figure 1 Sequence properties of the LCR16a duplication. **a**, Schematic display of the distribution pattern (red bars) of LCR16a duplications relative to a human chromosome 16 ideogram. The analysis is based on the published human genome project assembly^{4,5} and shows the clustering of duplications on the short arm of chromosome 16. **b**, The number of substitutions per site (K) among LCR16 duplications as a function of the number of aligned base pairs. Optimal global alignments for all possible pairwise combinations for each duplication were made and the degree of sequence identity for each alignment was computed ($n = 183$ pairwise alignments). LCR16a duplications are compared to all other characterized LCR16 duplications. (See Supplementary Information Fig. 1b for a detailed description of other duplications.) **c**, The gene structure of one member of the gene family (AF132984) is shown (green bars) compared with the 20-kb LCR16a segment from its corresponding genomic locus (AC002045). The analysis indicates eight exons, two strong polyadenylation signals within the 3' untranslated region, and a putative promoter region overlapping the first exon. PCR products used as probes in this study and their relative location are indicated above the gene structure.

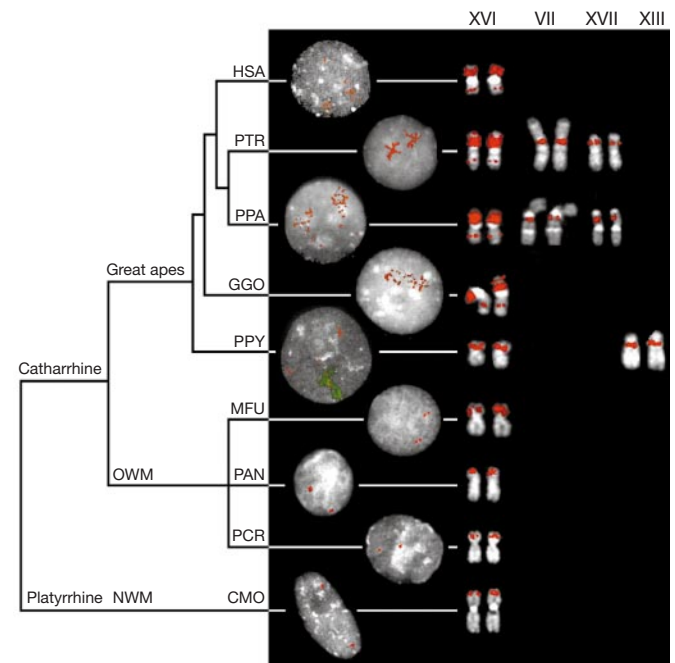


Figure 2 Comparative FISH analysis among primates. Metaphase (right) and interphase nuclei (left) that have been hybridized with probes (16.1/9 and 16.8/12) are shown from a representative panel of New World monkeys (NWM: CMO, *Callicebus mollochus*), Old World monkeys (OWM: MFU, *Macaca fascicularis*; PAN, *Papio anubis*; PCR, *Presbytis cristata*) and hominoid species (HSA, *Homo sapiens*; PTR, *Pan troglodytes*; PPA, *Pan paniscus*; GGO, *Gorilla gorilla*; PPY, *Pongo pygmaeus*). The results are depicted in the context of a generally accepted phylogeny of the species⁶. Roman numerals above metaphase chromosomes accord to standard cytogenetic nomenclature. Note that the multiple copies of the repeat located on XVI among the hominoids seem to paint the short arm of the chromosome. Reciprocal experiments using probes derived from other primate species were used to eliminate the possibility of false negative signal (see Methods). The orangutan (PPY) interphase also shows hybridization of a human chromosome XVI paint (green fluorescence). In this species, copies of the LCR16a duplication have spread to the pericentromeric region of chromosome XIII.

Species	HSA	PTR	HKL	PHA
HSA	–	0.021	0.034	0.074
PTR	0.002	–	0.038	0.080
HKL	0.004	0.004	–	0.070
PHA	0.008	0.008	0.007	–

HSA, *Homo sapiens* ($n = 14$ sequences); PTR, *Pan troglodytes* ($n = 17$); HKL, *Hylobates* ($n = 4$); and PHA, *Papio hamadryas* ($n = 1$); n is the number of paralogues analysed within each species. K is the average genetic distance (Kimura two-parameter model) between groups (above the diagonal); standard errors are given below the diagonal.

of approaches (interphase nuclei, library hybridization and sequence analysis of genomic clones), we estimated the copy number of the duplication in orangutans, gorillas, humans and chimpanzees as 9, 17, 15 and 25–30 copies, respectively. Interestingly, in both orangutans and chimpanzees, copies have been transposed to chromosomes other than 16 (Fig. 2), clearly indicating lineage-specific duplication events.

To more precisely estimate evolutionary timing of the duplications, we resequenced 1,421 base pairs (bp) of non-coding intronic sequence (Fig. 1c) from various human, chimpanzee, gibbon and baboon genomic subclones, and compared the number of nucleotide substitution events both within and between species (Table 1). First, we performed a Tajima's relative rate test using baboon sequence as an outgroup and orthologous pairs of sequence from chimpanzee, human and gibbon. True orthologues were determined by end-sequence analysis of the genomic subclones (see

Methods). This allowed duplicated copies that were orthologous by position to be identified within their respective genomes. On the basis of the analysis of four tests ($P = 0.109, 0.239, 0.242$ and 0.093), which indicated that the intronic sequence was evolving neutrally, we accepted the molecular-clock hypothesis. Using 25 Myr as an estimate of the timing of separation between humans and the Old World monkeys⁶, we calculated the mean rate of nucleotide substitution for intronic sequence as $(1.5 \pm 0.14) \times 10^{-9}$ substitutions per site per year. This estimate is remarkably consistent with previously published neutral rates between human and Old World monkeys ($1.2\text{--}1.8 \times 10^{-9}$) (ref. 7). On the basis of this rate of nucleotide substitution, we predict that the duplication events identified within the gibbons and orangutans occurred independently from those of humans and great apes. In the case of humans and chimpanzees, our analysis indicates that duplications occurred both before and after the separation of these two lineages (Fig. 3a). The observed quantitative and qualitative differences in the localization of some of the copies (Fig. 2) support this conclusion.

Alignments of the human paralogous segments revealed that regions corresponding to coding exons were conspicuously hyper-variable (10% nucleotide divergence when compared with intronic sequences that exhibited ~2% divergence). The increased frequency of substitution suggested rapid genic evolution had occurred along with the genomic dispersal of the LCR16a duplication. Increased substitutions among exons are a hallmark of genes undergoing adaptive evolution⁸⁻¹¹. A common test of positive selection is to

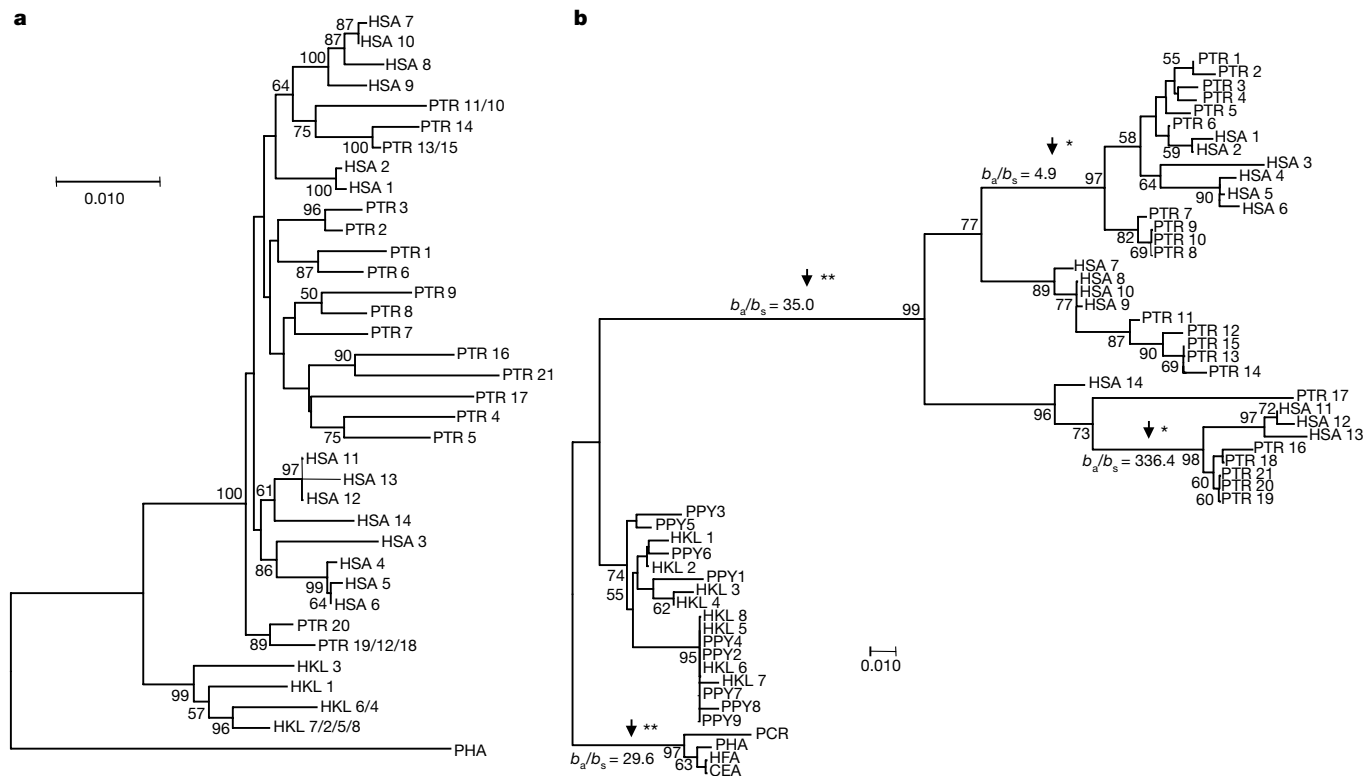


Figure 3 Phylogeny of coding and non-coding portions of the LCR16a duplication. Neighbour-joining phylogenetic trees for 1,421 bp of intronic sequence (**a**; introns 2, 3 and 4; Fig. 1c) and 186 bp of exon 2 (**b**) are compared. Extreme positive selection for exon 2 is indicated on the branch separating humans and African apes from the orangutan lineage (a 35-fold excess of amino-acid changes when compared with the neutral model). Note the significantly shorter branch lengths for flanking non-coding intronic sequences, which are consistent with nucleotide sites evolving at a neutral rate. More than 95% of the informative sites for the phylogenetic tree of exon 2 are the result of amino-acid-altering nucleotide changes. Branches showing significant positive selection are indicated by arrows with accompanying b/p quotients (estimated amino-acid replacement and

synonymous changes per branch per site¹²). Significance was calculated based on the difference (asterisk, $P < 0.05$; double asterisk, $P < 0.01$). Sequence for various duplicate copies are identified by species acronym (PHA, *Papio hamadryas*; HKL, *Hylobates klossi*; and see Fig. 2) and a number corresponding to clone and/or accession within GenBank (Supplementary Information Table 3). Scale bar, Jukes–Cantor corrected distance. The midpoint of all trees was set to one-half the distance between gibbon and baboon sequence taxa. Only bootstrap values $>50\%$ are shown ($n = 1,000$ replicates). A similar topology showing positive selection for exon 4 sequence was obtained (Supplementary Information Fig. 3).

compare the number of non-synonymous substitutions per site (K_a) to the number of synonymous substitutions per site (K_s)⁷. K_a/K_s quotients significantly greater than 1.0 are taken as evidence for positive selection. We assessed the K_a/K_s quotient for two of the most rapidly diverging exons (exons 2 and 4; Fig. 3b and Supplementary Information Fig. 3). Genomic subclones were obtained for the various copies of LCR16a from chimpanzees, gorillas, orangutans, gibbons and several Old World monkeys, and the exonic regions were comparatively sequenced (see Methods). Average K_a and K_s for all between-group and within-group comparisons was calculated independently for each exon (MEGA2, Modified Nei–Gojori method; Tables 2 and 3). A statistical test of the difference of average K_a and K_s values both between species and for multiple copies within species was used as a measure of significance.

In the case of exon 2, highly significant K_a/K_s quotients ($P < 0.0005$) were observed among all comparisons involving either humans or chimpanzees. The most extreme positive selection was observed between humans and Old World monkeys ($K_a/K_s = 13.0$, $P < 10^{-5}$) and between chimpanzees and Old World monkeys ($K_a/K_s = 11.8$, $P < 10^{-5}$). This level of amino-acid replacement translates into ~43% amino-acid divergence between these species and a rate of amino-acid replacement of $\sim 1.0 \times 10^{-8}$ changes per site per year for this exon. This is far in excess (20-fold) of most typical estimates⁷ of protein divergence between Old World and great ape species. Highly significant differences were also observed when comparing paralogues between chimpanzee and human sequences ($K_a/K_s = 5.0$, $P < 0.0001$) with an average amino-acid divergence of 23% among the paralogous exons. To identify more precisely when the major episode of positive selection occurred, we estimated the number of synonymous and non-synonymous nucleotide substitutions per site for each branch of a phylogenetic tree using the method first proposed by Zhang *et al.*¹². A major burst of positive selection seems to have occurred after the separation of the human and chimpanzee lineages from the orangutan (<12 Myr ago, $K_a/K_s = 35.0$), with subsequent protein-diversification events occurring during the emergence of chimpanzee and human species (see Fig. 3). A comparison with gorillas (Table 2) confirms that the major effect occurred in a common ancestor to humans and African apes. In stark contrast, the paralogues within orangutan and gibbon species have not experienced bursts of rapid positive selection (Table 2).

Similar to exon 2, analysis of exon 4 sequences showed a significant episode of positive selection after the separation of the chimpanzee/human and orangutan lineages ($K_a/K_s = 4.67$, $P < 0.05$;

Table 2 Positive selection of exon 2

Exon 2	$\bar{K}_a$ (s.e.)	$\bar{K}_s$ (s.e.)	$\bar{K}_a/\bar{K}_s$	$\bar{K}_a/\bar{K}_s$ (s.e.)	Z-value	P
HSA–PTR	0.189 (0.035)	0.042 (0.021)	4.5	0.147 (0.043)	3.42	<0.0001
HSA–GGO	0.176 (0.031)	0.066 (0.026)	2.67	0.110 (0.036)	3.06	<0.01
HSA–PPY	0.350 (0.065)	0.097 (0.048)	3.61	0.254 (0.080)	3.18	<0.0005
HSA–HKL	0.345 (0.062)	0.098 (0.045)	3.52	0.247 (0.077)	3.21	<0.0005
HSA–OW	0.429 (0.080)	0.033 (0.016)	13.0	0.396 (0.081)	4.89	<0.00001
PTR–GGO	0.182 (0.035)	0.069 (0.028)	2.64	0.114 (0.041)	2.78	<0.01
PTR–PPY	0.341 (0.064)	0.101 (0.048)	3.38	0.240 (0.077)	3.11	<0.0005
PTR–HKL	0.334 (0.062)	0.102 (0.046)	3.27	0.232 (0.075)	3.09	<0.01
PTR–OW	0.423 (0.078)	0.036 (0.016)	11.75	0.386 (0.078)	4.95	<0.00001
GGO–PPY	0.361 (0.067)	0.112 (0.048)	3.22	0.248 (0.083)	2.99	<0.01
GGO–HKL	0.350 (0.066)	0.113 (0.047)	3.10	0.244 (0.081)	2.77	<0.01
GGO–OW	0.420 (0.078)	0.054 (0.024)	7.78	0.366 (0.077)	4.75	<0.00001
PPY–HKL	0.025 (0.011)	0.038 (0.024)	0.66	-0.012 (0.020)	-0.55	NS
PPY–OW	0.113 (0.030)	0.071 (0.046)	1.59	0.042 (0.050)	0.84	NS
HKL–OW	0.105 (0.029)	0.072 (0.044)	1.46	0.033 (0.051)	0.65	NS
HSA–HSA	0.190 (0.032)	0.046 (0.030)	4.75	0.150 (0.041)	3.66	<0.001
PTR–PTR	0.181 (0.039)	0.046 (0.022)	3.93	0.135 (0.046)	2.93	<0.01
GGO–GGO	0.512 (0.031)	0.067 (0.030)	2.27	0.085 (0.037)	2.30	<0.01
PPY–PPY	0.033 (0.013)	0.037 (0.027)	0.89	-0.004 (0.023)	-0.17	NS
HKL–HKL	0.021 (0.012)	0.044 (0.028)	0.48	-0.026 (0.024)	-1.08	0.95
OW–OW	0.022 (0.012)	0 (0)	NA	0.022 (0.011)	2.00	<0.05

HSA, *Homo sapiens* ($n = 15$ paralogous sequences); PTR, *Par troglodytes* ($n = 21$); GGO, *Gorilla gorilla* ($n = 21$); HKL, *Hylobates klossi* ($n = 8$); PPY, *Pongo pygmaeus* ($n = 9$); OW, Old World monkeys (*Cercopithecus aethiops*, *Papio anubis*, *Papio hamadryas* and *Macaca fascicularis*; each species represented by a single sequence). NS, no significant positive selection detected.

Table 3 Positive selection of exon 4

Exon 4	$\bar{K}_a$ (s.e.)	$\bar{K}_s$ (s.e.)	$\bar{K}_a/\bar{K}_s$	$\bar{K}_a/\bar{K}_s$ (s.e.)	Z-value	P
HSA–PTR	0.099 (0.018)	0.056 (0.021)	1.77	0.043 (0.026)	1.65	<0.05
HSA–PPY	0.229 (0.043)	0.149 (0.053)	1.54	0.080 (0.066)	1.21	NS
HSA–HKL	0.233 (0.046)	0.143 (0.049)	1.63	0.090 (0.063)	1.36	NS
HSA–OW	0.275 (0.048)	0.177 (0.064)	1.55	0.098 (0.078)	1.26	NS
PTR–PPY	0.260 (0.048)	0.150 (0.054)	1.73	0.110 (0.068)	1.62	NS
PTR–HKL	0.261 (0.051)	0.143 (0.050)	1.83	0.118 (0.066)	1.79	<0.05
PTR–OW	0.292 (0.051)	0.180 (0.066)	1.62	0.112 (0.080)	1.40	NS
PPY–HKL	0.057 (0.020)	0.037 (0.021)	1.54	0.021 (0.024)	0.88	NS
PPY–OW	0.103 (0.025)	0.048 (0.023)	2.15	0.056 (0.033)	1.70	<0.05
HKL–OW	0.107 (0.026)	0.049 (0.025)	2.18	0.057 (0.032)	1.78	<0.05
HSA–HSA	0.078 (0.018)	0.066 (0.027)	1.18	0.012 (0.035)	0.34	NS
PTR–PTR	0.089 (0.015)	0.045 (0.016)	1.98	0.044 (0.019)	2.32	<0.01
PPY–PPY	0.060 (0.019)	0.039 (0.025)	1.54	0.021 (0.028)	0.75	NS
HKL–HKL	0.066 (0.021)	0.040 (0.025)	1.65	0.026 (0.028)	0.93	NS
OW–OW	0.093 (0.020)	0.055 (0.028)	1.69	0.039 (0.032)	1.22	NS

HSA, *Homo sapiens* ($n = 15$ paralogous sequences); PTR, *Par troglodytes* ($n = 19$); HKL, *Hylobates klossi* ($n = 7$); PPY, *Pongo pygmaeus* ($n = 14$); OW, Old World monkeys (*Cercopithecus aethiops*, *Papio anubis*, *Papio hamadryas*, *Presbytis Cristata* and *Allenopithecus*, each species represented by a single copy sequence). NS, no significant positive selection detected.

Supplementary Information Fig. 3). Although there is a marked increase in the number of putative amino-acid replacements (~30% between chimpanzees/humans and Old World monkeys), much more modest K_a/K_s quotients (1.81–2.45, Table 3) are observed in comparisons between and within species. This reduction is primarily due to the greater number of synonymous events that have occurred concurrently with non-synonymous changes. These events have occurred precisely within the putative coding regions of the exons and do not extend into flanking intronic sequences. Trivial explanations for the enhanced rates of non-synonymous and synonymous changes were examined, including unusual CpG (cytosine–guanine) content, codon bias and the presence of hypermutable repeat sequences within the exonic regions. No evidence in support of these alternatives could be found. It should be noted, however, that alternative splicing has been observed for exon 4 among human complementary DNAs (for example AF229069, D86974 and Supplementary Information Fig. 2). It is possible that some paralogues in different species have also experienced alternative splicing, resulting in new protein products with open reading frames that no longer conform to that predicted by our human reference cDNA (AF132984). The result of such an apparent frameshift would be to increase both synonymous and non-synonymous rates if both splice variants were represented in each species. Furthermore, no distinction in this study has been made between functional and non-functional paralogues, because this would require detailed expression and protein analyses. We felt that this treatment was conservative as pseudogene comparisons and alternative splicing would tend to neutralize both adaptive and purifying selection constraints. Consequently, such events would cause K_a/K_s quotients to approximate unity and reduce our power to detect positive selection.

Although the precise function of this gene family is unknown, previous examples of positive selection have included either genes involved in xenobiotic recognition of macromolecules (immunoglobulin genes, venom toxins, lysozymes)^{10,12–14} or genes associated with male reproduction^{8,9,11,15}. In many of these cases, positive selection has occurred in concert with duplication events. Delineation of the function of this gene family will require detailed experimental analysis. In humans, multiple transcripts ($n = 284$) with open reading frames have been recovered, demonstrating clear transcriptional and splicing potency. Analysis by polymerase chain reaction with reverse transcription (RT-PCR) confirms a broad distribution of this gene family in most human tissues (Supplementary Information Fig. 4). Finally, immunolocalization studies performed with constructs fused with green fluorescent protein (GFP) reveal a clear localization to the nuclear membrane for at least one translated member of this gene family. Colocalization of these products with antibodies raised against membrane-bound nucleo-

porin (p62)¹⁶ further indicates that this particular human copy may associate with the nuclear pore complex (Supplementary Information Fig. 5).

Our analysis has revealed an extraordinary degree of evolutionary plasticity, at the level of both the genome and the gene. We provide evidence for the evolution of a hominoid gene family by recent duplication and positive selection. Can additional examples within the human proteome be expected? Preliminary analysis of the human genome suggests that as much as 5–7% of all human sequences may have been duplicated within the last 30 Myr of evolution. The abundance of segmental duplications may be an important reservoir for the emergence of other hominoid genes that do not possess definitive orthologues in the genomes of model organisms. □

Methods

Human genome analysis

Searches for sequence similarity in GenBank (release 121.0) identified 41 human accessions that contained a complete copy of the LCR16a repeat. Because the degree of sequence similarity among these copies approached levels of allelic variation (98–99%), comparison of unique sequences, flanking the duplications, and partial sequencing of chromosome 16 cosmids (LA16NC02) were used to distinguish various paralogues from allelic overlap. The human cosmid library is derived from a single chromosome 16 haplotype, thereby allowing sequence variants to effectively classify duplicated copies¹⁷. A suite of genomic software tools were used to analyse and characterize the duplications, including PARASIGHT³ to delineate the junction sequences and the extent of overlap for each duplicated segment, ALIGN to perform optimal global pairwise alignments between copies, and sim4 to optimally compare cDNA with genomic DNA¹⁷. Only pairwise sequence alignments greater than 1 kb with at least 90% identity were considered in this analysis (see Supplementary Information). Unique sequence differences within the predicted exons from genomic sequences compared with expressed sequence tag sequences were used to identify transcriptionally competent loci. Software for analysis of protein structure (<http://bmerc-www.bu.edu/psa> and <http://maple.bioc.columbia.edu/predictprotein>) predicted the presence of single transmembrane domain flanked by α -helical secondary structure for the AF132984 open reading frame (347 amino acids).

Fluorescence *in situ* hybridization

Chromosome metaphase and interphase nuclei were prepared from lymphoblastoid cell lines representative of five hominoid species (*Homo sapiens*, *Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla* and *Pongo pygmaeus*), three Old World monkey (*Pan anubis*, *Presbytis cristata* and *Cercopithecus aethiops*) and one New World monkey (*Callicebus mollochus*). *In situ* hybridizations were performed under standard conditions¹⁸ with two genomic probes (16.1/9 and 16.8/12; Fig. 1c) subcloned from one of the paralogous copies (AC002039). To eliminate the effect of cross-hybridization of common repeat sequences, probes were blocked by using repetitive DNA (C_{α}) before hybridization. At least 20 independent metaphase and interphase nuclei were examined in the determination of copy number and chromosomal band location. The combined probes spanned 11.5 kb of genomic sequence and included 7 exons of the AF132984 cDNA. Reciprocal experiments using probes derived from baboons and gibbons were used to confirm the specificity of hybridization. When necessary, hybridizations were performed in conjunction with human whole-chromosome painting probes to confirm chromosomal assignment (orangutan, Fig. 2).

Library hybridization and sequencing

Large-insert genomic libraries from human (LA16NC02), chimpanzee (RPCI-43, *P. troglodytes*), gibbon (DKZ-140, *Hylobates klossi*) and the olive baboon (RPCI-41, *Papio hamadryas*) were hybridized with PCR-amplified products (16.1/9 and 16.8/12; see Supplementary Information Table 3 for conditions and oligonucleotide sequences). All hybridizations were performed as previously described¹⁷. A total of 156 genomic clones (70 human, 75 chimpanzee, 10 gibbon and 1 baboon) were comparatively sequenced. (In all, 1,753 bp were examined, partitioned into 1,421 bp of intronic sequence and 332 bp of sequence from exons 2 and 4.) All PCR products (forward and reverse reactions) were directly sequenced using a modified dye-terminator sequencing protocol¹⁷. Non-human sequences were deemed to be paralogous if more than two sequence differences were observed within 150 bp of coding sequence. All paralogues were encoded by species name and numbered according to clone and/or accession identifier (Supplementary Information Table 4). End sequences generated from the cloning site (T7 and T3 or T3 and SP6) were used to further position specific paralogous copies with respect to the human genome reference. In all cases, the duplicated sequence was flanked either directly or within ~70 kb by non-duplicated unique sequence. As a result, end-sequence analysis allows a subset of bacterial artificial chromosome (BAC) clones from different species to be unambiguously placed on the basis of alignment to unique sequence on either side spanning the duplication. For exons 2 and 4, additional sequences were generated by TA-subcloning of PCR-amplified product from orangutans (*P. pygmaeus*) and direct PCR sequencing of products from various Old World monkeys (*Macaca fascicularis*, *Presbytis cristata* and *Cercopithecus aethiops*).

Sequence analysis

Estimates of genetic distance (pairwise deletion) were calculated using the Jukes and Cantor one-parameter model (when transition/transversion quotients $i/v \approx 1.0$) or Kimura's two-parameter model (when $i/v \approx 2.0$)¹⁹. A Tajima's relative rate test was performed²⁰ using orthologous sequence pairs (HSA13 versus PTR3, PTR8 versus HKL1, HKL1 versus HSA13, and HSA3 versus PTR17; see Supplementary Information Table 4) from human, chimpanzee and gibbon intronic sequences with the baboon sequence (PHA) as the outgroup. Four such tests were used to accept the molecular-clock hypothesis for the non-coding sequences under study in this analysis. Estimates of duplication timing were based on 1,421 bp of non-coding sequence and were calculated using the formula $r = K/2T$, (where r is the rate of nucleotide substitution; K is the number of substitutions for site; and T is the time of separation), with the baboon sequence as a reference orthologue and an estimated time of separation from the hominoid lineage of 25 Myr⁶. For exonic sequence, the average number of synonymous (K_s) and non-synonymous (K_a) substitutions per site were estimated using the modified Nei–Gojobori method^{12,21}. To test for positive darwinian selection, we calculated the difference ($D = K_a - K_s$) within and between groups for all pairwise comparisons of paralogues. Groups are here defined as species. Owing to the large number of pairwise analyses performed, significance levels should be corrected for multiple comparisons. Because the comparisons are not independent of one another, the usual Bonferroni method cannot be used. Instead, the average difference for all comparisons and its associated standard error were computed. Initially, all possible comparisons were made among the sequenced exons and the difference between amino-acid and synonymous substitutions was calculated for each pairwise comparison. The differences were averaged between and within groups (within groups included multiple duplicate copies within each species). The variance for the average difference was estimated using the bootstrap method ($n = 1,000$ replicates) and a one-tailed Z-test ($Z = D/\sigma$) to determine the level of significance²². Positive selection was defined as a significant positive difference. Evolutionary trees of multiple aligned sequences (ClustalW) were generated using neighbour-joining distance estimates (MEGA2). Only bootstrap values >50% are indicated in the tree topology. Internal branch estimates of the number of synonymous (b_s) and non-synonymous (b_a) substitutions per site were determined by the method of Zhang *et al.*¹².

Received 13 March; accepted 29 June 2001.

- Ohno, S. *Evolution by Gene Duplication* (Springer, Berlin, 1970).
- Stallings, R., Whitmore, S., Doggett, N. & Callen, D. Refined physical mapping of chromosome 16-specific low-abundance repetitive DNA sequences. *Cytogenet. Cell Genet.* **63**, 97–101 (1993).
- Loftus, B. *et al.* Genome duplications and other features in 12 Mbp of DNA sequence from human chromosome 16p and 16q. *Genomics* **60**, 295–308 (1999).
- The International Human Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–920 (2001).
- Bailey, J. A., Yavor, A. M., Massa, H. F., Trask, B. J. & Eichler, E. E. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* **11**, 1005–1017 (2001).
- Goodman, M. The genomic record of humankind's evolutionary roots. *Am. J. Hum. Genet.* **64**, 31–39 (1999).
- Li, W. *Molecular Evolution* (Sinauer, Sunderland, 1997).
- Wyckoff, G. J., Wang, W. & Wu, C. I. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403**, 304–309 (2000).
- Nurminsky, D. I., Nurminskaya, M. V., De Aguiar, D. & Hartl, D. L. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature* **396**, 572–575 (1998).
- Duda, T. F. & Palumbi, S. R. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc. Natl Acad. Sci. USA* **96**, 6820–6823 (1999).
- Vaquier, V. D., Swanson, W. J. & Lee, Y. H. Positive Darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence? *J. Mol. Evol.* **44**, S15–S22 (1997).
- Zhang, J., Rosenberg, H. F. & Nei, M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl Acad. Sci. USA* **95**, 3708–3713 (1998).
- Hughes, A. L. & Nei, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167–170 (1988).
- Messier, W. & Stewart, C. B. Episodic adaptive evolution of primate lysozymes. *Nature* **385**, 151–154 (1997).
- Ting, C. T., Tsaur, S. C., Wu, M. L. & Wu, C. I. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**, 1501–1504 (1998).
- Davis, L. I. & Blobel, G. Identification and characterization of a nuclear pore complex protein. *Cell* **45**, 699–709 (1986).
- Horvath, J., Schwartz, S. & Eichler, E. The mosaic structure of a 2p11 pericentromeric segment: A strategy for characterizing complex regions of the human genome. *Genome Res.* **10**, 839–852 (2000).
- Lichter, P. *et al.* High-resolution mapping of human chromosome 11 by *in situ* hybridization with cosmid clones. *Science* **247**, 64–69 (1990).
- Jukes, T. H. & Cantor, C. R. in *Mammalian Protein Metabolism* (ed. Munro, H. N.) 21–123 (Academic, New York, 1969).
- Tajima, F. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**, 599–607 (1993).
- Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426 (1986).
- Nei, M. & Kumar, S. *Molecular Evolution and Phylogenetics* (Oxford Univ. Press, New York, 2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank W. E. Kutz and D. Zivkovic for technical assistance and sequencing analyses. This work was supported by grants from the National Institutes of Health and the US Department of Energy to E.E.E., and grants from Progetti di Interesse Nazionale (PRIN), Centro Eccellenza (CE), Ministero per la Ricerca Scientifica e Tecnologica (MURST) and Telethon to M.R. We are grateful to C. I. Wu, A. Chakravarti, D. Cutler, D. Locke, G. Matera and H. Willard for comments on this manuscript.

Correspondence and requests for materials should be addressed to E.E.E. (e-mail: eee@po.cwru.edu). All sequences have been deposited in GenBank under accession numbers AF364182–AF364299.

A forkhead-domain gene is mutated in a severe speech and language disorder

Cecilia S. L. Lai^{*†}, Simon E. Fisher^{*†}, Jane A. Hurst[‡], Faraneh Vargha-Khadem[§] & Anthony P. Monaco^{*}

^{*} Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

[‡] Department of Clinical Genetics, Oxford Radcliffe Hospital, Oxford OX3 7LJ, UK

[§] Developmental Cognitive Neuroscience Unit, Institute of Child Health, Mecklenburgh Square, London WC1N 2AP, UK

[†] These authors contributed equally to this work

Individuals affected with developmental disorders of speech and language have substantial difficulty acquiring expressive and/or receptive language in the absence of any profound sensory or neurological impairment and despite adequate intelligence and opportunity¹. Although studies of twins consistently indicate that a significant genetic component is involved^{1–3}, most families segregating speech and language deficits show complex patterns of inheritance, and a gene that predisposes individuals to such disorders has not been identified. We have studied a unique three-generation pedigree, KE, in which a severe speech and language disorder is transmitted as an autosomal-dominant monogenic trait⁴. Our previous work mapped the locus responsible, SPCH1, to a 5.6-cM interval of region 7q31 on chromosome 7 (ref. 5). We also identified an unrelated individual, CS, in whom speech and language impairment is associated with a chromosomal translocation involving the SPCH1 interval⁶. Here we show that the gene *FOXP2*, which encodes a putative transcription factor containing a polyglutamine tract and a forkhead DNA-binding domain, is directly disrupted by the translocation breakpoint in CS. In addition, we identify a point mutation in affected members of the KE family that alters an invariant amino-acid residue in the forkhead domain. Our findings suggest that *FOXP2* is involved in the developmental process that culminates in speech and language.

Investigations of the KE family (Fig. 1) have been central to discussions regarding the innate aspects of language ability^{4,5,7–9}. Affected members have a severe impairment in the selection and sequencing of fine orofacial movements, which are necessary for articulation (referred to as a developmental verbal dyspraxia; MIM 602081)^{4,8,9}. The disorder is also characterized by deficits in several facets of language processing (such as the ability to break up words into their constituent phonemes) and grammatical skills (including production and comprehension of word inflections and syntactical structure)^{7,8}.

Although the mean non-verbal IQ of affected members is lower than that of unaffected members⁸, there are affected individuals in the family who have non-verbal ability close to the population

average, despite having severe speech and language difficulties; therefore, non-verbal deficits cannot be considered as characteristic of the disorder. Functional and structural brain-imaging studies of affected members of the KE family have suggested that the basal ganglia may be a site of bilateral pathology associated with the trait⁹. Although there has been some debate over which feature of the phenotype constitutes the core deficit in this disorder, all the different studies agree that the gene disrupted in the KE family is likely to be important in neural mechanisms mediating the development of speech and language.

After our initial localization of SPCH1 to 7q31 (ref. 5), we used a bioinformatic approach to construct a transcript map of the crucial interval containing nearly 8 megabases of completed genomic sequence⁶. In addition, we reported molecular cytogenetic studies of an unrelated patient CS, who has a speech and language disorder that is strikingly similar to that of the KE family, associated with a *de novo* balanced reciprocal translocation t(5;7)(q22;q31.2)⁶. As observed for affected members of the KE family, CS presents with a severe orofacial dyspraxia despite normal early feeding and gross motor development. For both KE and CS phenotypes, there is substantial impairment of expressive and receptive language abilities. In both cases, general intelligence is relatively spared: although there is some lowering of IQ, deficits are more profound in the verbal domain.

Fluorescence *in-situ* hybridization (FISH) with a series of bacterial artificial chromosome (BAC) clones enabled us to map the 7q31.2 breakpoint of CS to a single clone, named NH0563O05, and did not reveal any additional associated genomic rearrangements in the vicinity of the translocation⁶. We discovered that the NH0563O05 clone contains several exons from CAGH44, a brain-expressed transcript encoding a large stretch of consecutive polyglutamines⁶ (Fig. 2). A previous study of CAGH44 had determined only the first 869 base pairs (bp) of coding sequence from a partial transcript of the gene, in which no in-frame stop codon had been reached¹⁰. Investigation of this 5' part of the open reading frame (ORF) in the KE family did not detect any sequence variant co-segregating with the speech and language disorder⁶.

To isolate the complete coding region of this candidate gene, we obtained the genomic sequence of NH0563O05 and adjacent BAC clones. Computer-based investigation of these data, using database search tools and gene prediction programs, enabled us to assemble the sequence of a hypothetical 2.5-kilobase (kb) transcript comprising 17 exons and containing a complete ORF of about 2.1 kb (Fig. 2). We verified the predicted transcript sequence experimentally (see Methods), confirming the exon–intron structure of the gene and identifying alternative splicing of two additional exons at the 5' end of the gene in all tissues examined (Fig. 2b). The carboxy-terminal portion of the predicted protein sequence encoded by this gene

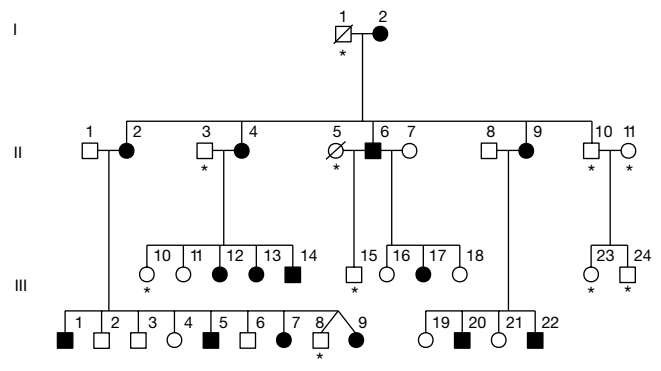


Figure 1 Pedigree of the KE family. Affected individuals are indicated by filled symbols. Asterisks indicate those individuals who were unavailable for genetic analyses. Squares are males, circles are females, and a line through a symbol indicates that the person is deceased.

contains a segment of 84 amino acids (encoded by exons 12–14) that shows high similarity to the characteristic DNA-binding domain of the forkhead/winged-helix (FOX) family of transcription factors^{11–14} (Fig. 2c). The complete gene has been therefore designated FOXP2, in accordance with the standard nomenclature proposed for this rapidly growing gene family¹⁴.

Northern blot analysis (see Supplementary Information) of several human adult tissues showed that there is broad expression of a roughly 6.5-kb transcript. This transcript was also observed in fetal tissues, with strong expression in brain. Similarly, an investigation of the murine homologue of FOXP2 has demonstrated expression in a range of adult and fetal mouse tissues¹⁵. Using *in situ* hybridization, it was also found that murine FOXP2 is expressed in

defined regions of the central nervous system during mouse embryogenesis, including the neopallial cortex and the developing cerebral hemispheres¹⁵.

We used additional FISH experiments and Southern blot analysis of DNA from CS to investigate further the relationship between the translocation and the FOXP2 locus. We thereby localized the translocation breakpoint to a 200-bp region in the intron between exons 3b and 4 (Fig. 3). These results indicate that disruption of FOXP2 is implicated in the aetiology of the speech and language disorder of this patient.

We screened the newly defined coding regions (exons 1, 3b and 8–17) of FOXP2 for mutations in the KE family. A G-to-A nucleotide transition was detected in exon 14 of affected individuals, and

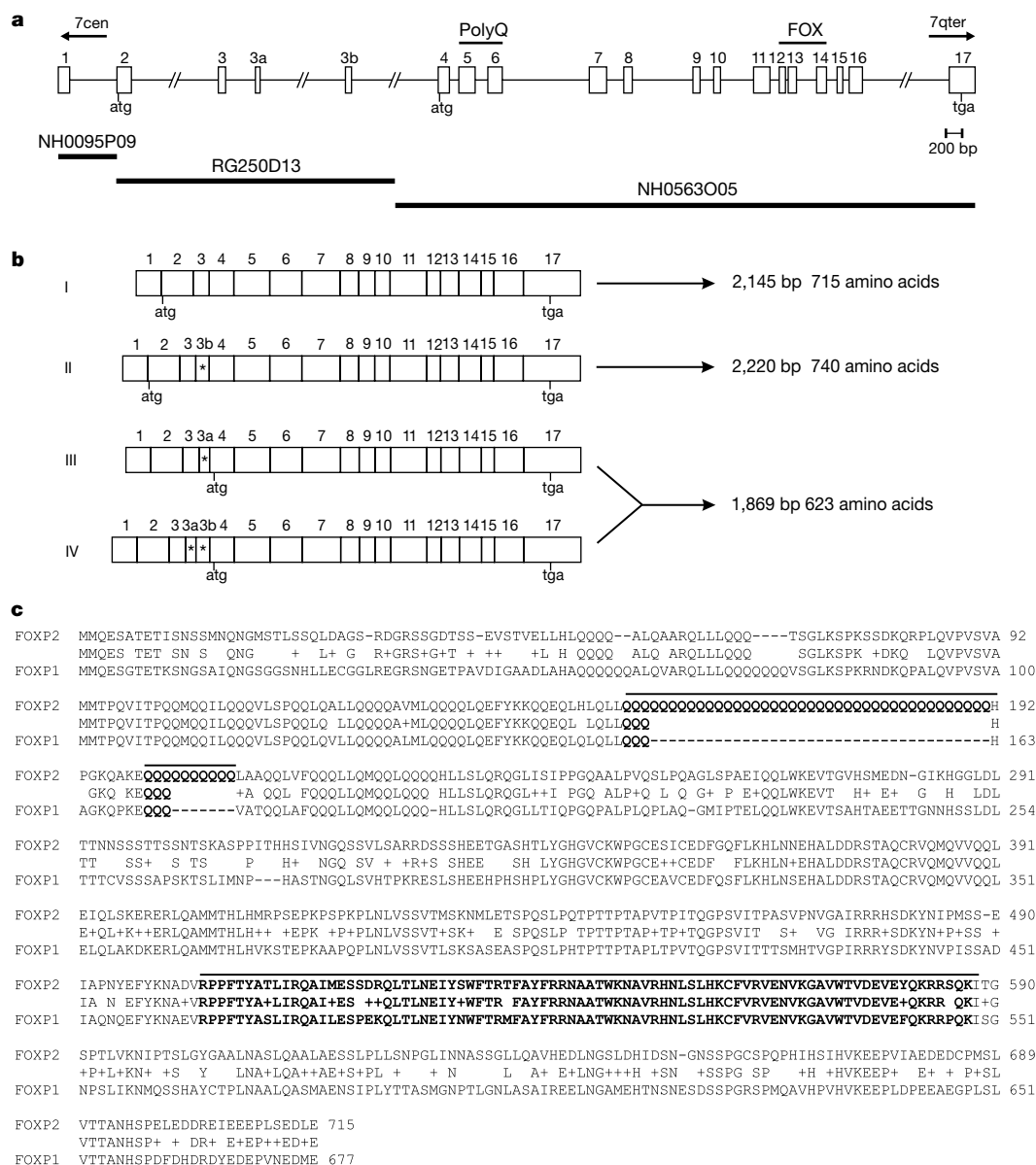


Figure 2 Identification of the human FOXP2 gene. **a**, Representation of human FOXP2 gene structure. Boxes represent exons, with positions of initiation and termination codons indicated. The scale shown applies only to exons; the entire region spans more than 267 kb of genomic DNA. Exons encoding polyglutamine tracts (PolyQ) and the forkhead domain (FOX) are indicated. The gene includes regions corresponding to expressed sequence tag ye5f03.r1 (exon 1), the partial CAGH44 transcript (exons 2–7) and a partial cDNA clone YX52E07 (exons 11–15). BAC genomic sequence entries are aligned beneath the gene structure. **b**, Alternative splicing of exons 3a and 3b (indicated by asterisks) leads to four different transcripts. 'I' was originally identified by

genomic predictions. 'II' contains exon 3b, which inserts 75 bp in-frame into the coding region. 'III' and 'IV' include the 58-bp exon 3a, which shifts the frame such that the ORF begins in exon 4, rather than exon 2. **c**, Amino-acid sequence encoded by human FOXP2 (transcript 'I'), aligned with human FOXP1 (accession AAG47632). The 40-residue and 10-residue stretches of polyglutamine in FOXP2 are reduced to only three glutamines each in FOXP1. FOXP2 transcript 'II' inserts 25 amino acids between residues 86 and 87. Transcripts 'III' and 'IV' give a shorter product beginning with the methionine at

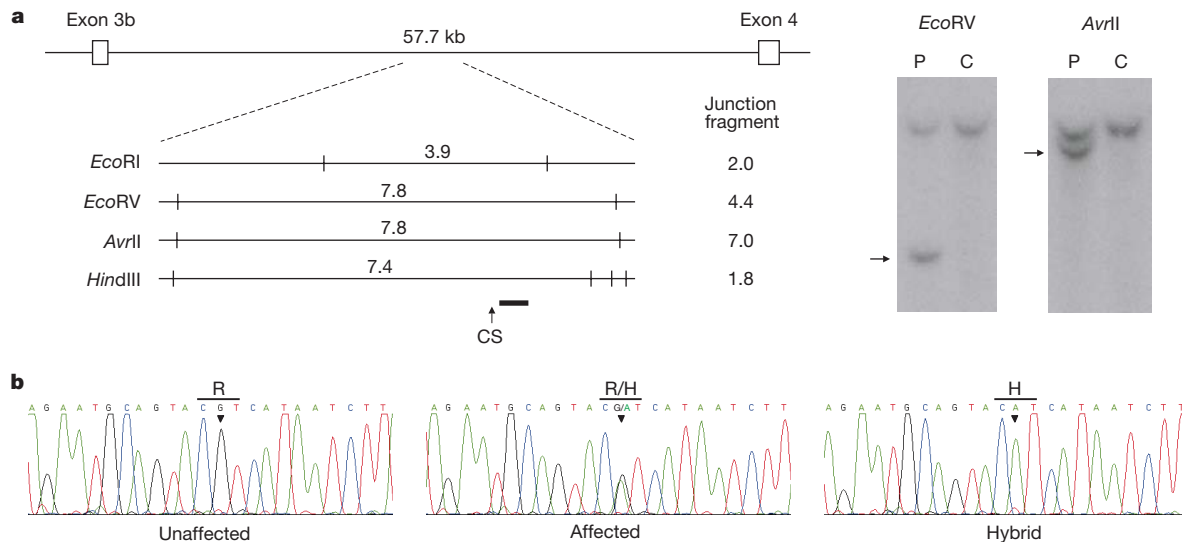


Figure 3 Disruption of FOXP2 in patients with severe speech and language disorder. **a**, Localization of the CS translocation breakpoint. A 499-bp probe (indicated by a thick black line) from the intron between exon 3b and exon 4 detected abnormal restriction fragments on Southern blots with four different enzymes, *EcoRI*, *EcoRV*, *AvrII* and *HindIII*. A scaled restriction map of the normal locus is shown, with estimated sizes (in kilobases) of detected junction fragments displayed at the side. The *HindIII* results indicate that the CS breakpoint maps to a region of ~200 bp on the centromeric side of the probe. Examples of Southern blot hybridizations with digested DNA from the patient (P) and a

control (C) are also shown for two of the enzymes, with the junction fragments indicated by arrows. **b**, Direct sequencing of exons from FOXP2 detected a G-to-A transition causing an R553H substitution in the forkhead domain in family KE. All affected individuals from the KE pedigree were heterozygous for this mutation, whereas all unaffected individuals were homozygous for the wild type (see Fig. 1). Somatic cell hybrids containing only the chromosome 7 associated with the speech and language disorder⁶ were hemizygous for the mutation.

shown to co-segregate perfectly with the speech and language disorder in the KE pedigree (Fig. 3). Using a restriction-enzyme-based assay, we showed that the mutation was absent in 364 independent chromosomes from normal Caucasian controls (data not shown), indicating that it does not represent a naturally occurring polymorphism. The mutation is predicted to result in an arginine-to-histidine substitution (R553H) in the forkhead DNA-binding domain of FOXP2 (Fig. 4). Forkhead (or winged-helix) domains adopt a characteristic structure, comprising three amphipathic α -helices followed by two large loops (called 'wings'), in which the third α -helix is presented to the major groove of the target DNA^{12,16}. The R553H change occurs in this third helix, which is the most highly conserved part of the forkhead domain¹², adjacent to a histidine residue that makes a direct base contact with the target DNA¹⁶.

The R553 amino acid is invariant in all the currently known

members of the large family of forkhead proteins, in species ranging from yeast to human (see <http://www.biology.pomona.edu/fox.html>). Furthermore, it has been proposed as an invariant feature of all homeodomain recognition helices¹². Therefore, we suggest that this arginine residue is crucially important for the function of the forkhead domain, and that the histidine substitution observed in affected members of the KE family disrupts the DNA-binding and/or transactivation properties of FOXP2. The alternative hypothesis—that the R553H change is in linkage disequilibrium with a pathogenic mutation in a neighbouring gene and that the disorder in the translocation patient actually results from positional inactivation of this other gene—is highly unlikely.

Many members of the forkhead family are known to be key regulators of embryogenesis¹³. Mutations in FOX genes have been implicated in specific human disorders, including congenital

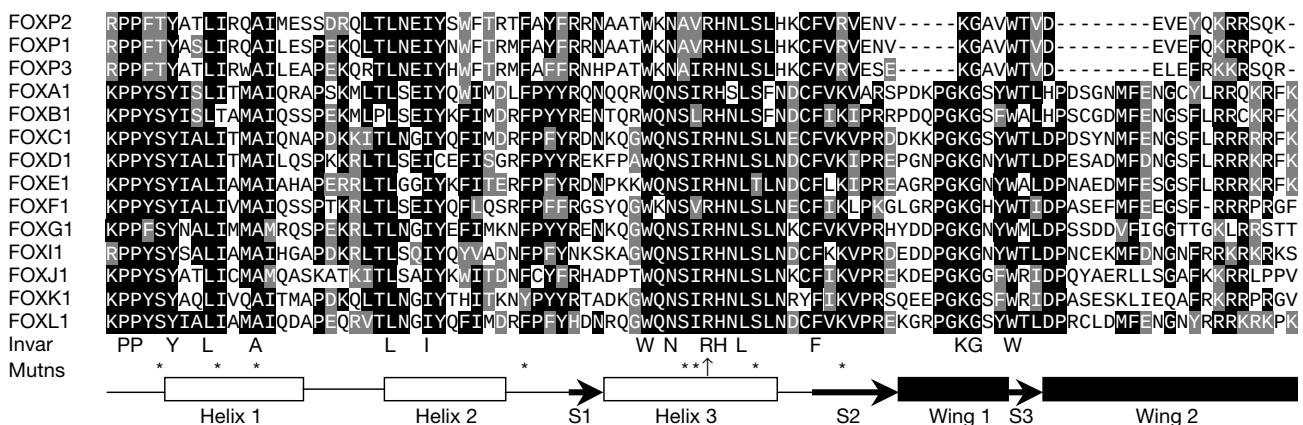


Figure 4 Forkhead domains of the three known FOXP proteins aligned with representative proteins from several branches of the FOX family. All sequences are from *Homo sapiens*. Residues that are invariant in this selection of forkhead proteins are given beneath the alignment. Asterisks show sites of the substitution mutations in FOXC1, FOXC2 and FOXP2 that have been previously implicated in human disease states^{17–19,23,24}. The upwards

arrow indicates the site of the R553H substitution identified in FOXP2 in affected members of the KE pedigree. The proposed structure of the forkhead domain as established by X-ray crystallography¹⁶ is shown, containing three α -helices, three β -strands (S1–3) and

glaucoma (FOXC1)^{17,18}, thyroid agenesis (FOXE1)¹⁹, lymphedema-distichiasis (LD) syndrome (FOXC2)²⁰, blepharophimosis/ptosis/epicanthus inversus (BPES) syndrome (FOXL2)²¹, and anterior-segment dysgenesis associated with cataracts (FOXE3)²². The mouse phenotype *scuffy* and a similar syndrome found in humans both result from disruption of FOXP3 (refs 23–25), a gene that is closely related to FOXP2.

A significant number of the mutations identified in FOX genes are missense changes, and all of these result in substitution at residues in the forkhead domain^{17–19,23,24}, as observed here for FOXP2 (Fig. 4). Frameshift and nonsense mutations yielding truncated protein products that lack a forkhead domain have also been identified^{17,18,20,23}. In addition, there have been reports of balanced translocations causing positional effect inactivation of FOXC1, FOXC2 and FOXL2 in glaucoma¹⁷, lymphedema-distichiasis²⁰ and BPES²¹, respectively. Data from those studies^{17–20,23,24}, as well as from mouse models^{25–27} and *in vitro* functional assays¹⁹, indicate that inactivation or loss of the forkhead domain is a general mechanism by which mutation of FOX genes can lead to human disease states. Investigations of forkhead-domain mutations associated with autosomal dominant traits suggests that the resulting disorders are a consequence of haplo-insufficiency during embryological development^{17,18,20,27}. The finding that duplications involving FOXC1 can cause anterior-chamber defects of the eye^{28,29} provides further evidence that the correct gene dosage of forkhead transcription factors is important in embryogenesis.

In addition to the forkhead domain, the FOXP2 protein also contains a stretch of 40 consecutive glutamines followed by a second stretch of only 10 glutamines. Abnormal expansion of variable polyglutamine tracts has been implicated in several hereditary neurodegenerative disorders³⁰. The polyglutamine region of FOXP2 is encoded by a mixture of CAG and CAA codons, making it highly stable in normal individuals¹⁰. Although polyglutamine tracts have been found in many transcription-related proteins³⁰ this is the first report of such a domain in a FOX family member. The amino-acid sequence of FOXP2 shows remarkable similarity throughout its length to FOXP1, another member of the P branch of the forkhead family that has been identified in humans (68% identity; 80% similarity). However, an intriguing difference between these two human paralogues is that the polyglutamine tracts of FOXP2 are reduced markedly in FOXP1 (Fig. 2c); thus, comparison of the properties of the two proteins might shed light on the role of polyglutamine repeats in non-pathological processes.

In conclusion, we have shown that the FOXP2 gene is directly disrupted by a translocation in a patient with a speech and language disorder, and that a mutation affecting a crucial residue of the forkhead domain of this putative transcription factor co-segregates with affection status in the KE family. We propose that, in both cases, FOXP2 haplo-insufficiency in the brain at a key stage of embryogenesis leads to abnormal development of neural structures that are important for speech and language. This is the first gene, to our knowledge, to have been implicated in such pathways, and it promises to offer insights into the molecular processes mediating this uniquely human trait. □

Methods

Bioinformatic analyses

We obtained BAC genomic sequence data from the Washington University Genome Sequencing Center database (<http://genome.wustl.edu/gsc/>). Genomic sequence data were analysed with database search tools and gene prediction software, as implemented in the NIX package (<http://www.hgmp.mrc.ac.uk/NIX/>). Amino-acid sequences of FOXP2 and FOXP1 in Fig. 2c were aligned using BLAST2 (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html>). Forkhead-domain sequences from human FOX proteins in Fig. 4 were aligned using ClustalW, accessed through the Baylor College of Medicine Search Launcher (<http://searchlauncher.bcm.tmc.edu:9331/multi-align/multi-align.html>).

FOXP2 mRNA sequence and genomic structure

We used a reverse-transcriptase polymerase chain reaction (RT-PCR)-based approach to

confirm the FOXP2 mRNA sequence that had been predicted by bioinformatics. Primers were designed from putative exonic sequence and used to amplify by PCR first-strand complementary DNA from a range of adult tissues, which was obtained from Clontech. Products were sequenced as described⁶ and compared with the predicted sequence.

Expression analyses of FOXP2

Adult and fetal northern blots were obtained from Clontech and hybridized according to the manufacturers' instructions, using a cDNA probe isolated from exons 8–11 of FOXP2.

Translocation mapping

We performed FISH on metaphase spreads of cells from CS, using a series of roughly 10-kb genomic probes obtained from the NH0563O05 BAC clone, as described⁶. In parallel, we ran Southern blot analyses of several restriction fragments spanning the FOXP2 locus, comparing digested DNA from CS with that from unaffected controls, according to standard procedures.

Mutation search

On the basis of genomic sequence information, we designed primers to flank each FOXP2 exon. These were used for PCR amplification of DNA from affected and unaffected individuals of the KE family, and from hybrid cell lines containing the affected chromosome 7 (ref. 6). We sequenced products as described⁶. The G-to-A transition detected in exon 14 of affected individuals destroys a restriction site for the enzyme *Mae*II (A|CGGT). An assay using this restriction enzyme was developed to test for the exon 14 change in 182 unrelated normal controls.

GenBank accession numbers

BAC genomic sequence data, AC073626, AC003992 and AC020606; human FOXP2 mRNA sequence, AF337817.

Received 13 February; accepted 27 July 2001.

- Bishop, D. V. M., North, T. & Donlan, C. Genetic basis for specific language impairment: evidence from a twin study. *Dev. Med. Child Neurol.* **37**, 56–71 (1995).
- Tombli, J. B. & Buckwalter, P. R. Heritability of poor language achievement among twins. *J. Speech Lang. Hear. Res.* **41**, 188–199 (1998).
- Dale, P. S. *et al.* Genetic influence on language delay in two-year-old children. *Nature Neurosci.* **1**, 324–328 (1998).
- Hurst, J. A., Baraitser, M., Auger, E., Graham, F. & Norell, S. An extended family with a dominantly inherited speech disorder. *Dev. Med. Child Neurol.* **32**, 347–355 (1990).
- Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P. & Pembrey, M. E. Localization of a gene implicated in a severe speech and language disorder. *Nature Genet.* **18**, 168–170 (1998).
- Lai, C. S. *et al.* The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. *Am. J. Hum. Genet.* **67**, 357–368 (2000).
- Gopnik, M. & Crago, M. B. Familial aggregation of a developmental language disorder. *Cognition* **39**, 1–50 (1991).
- Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P. & Passingham, R. Pragmatic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proc. Natl Acad. Sci. USA* **92**, 930–933 (1995).
- Vargha-Khadem, F. *et al.* Neural basis of an inherited speech and language disorder. *Proc. Natl Acad. Sci. USA* **95**, 12695–12700 (1998).
- Margolis, R. L. *et al.* cDNAs with long CAG trinucleotide repeats from human brain. *Hum. Genet.* **100**, 114–122 (1997).
- Lai, E., Clark, K. L., Burley, S. K. & Darnell, J. E. Jr Hepatocyte nuclear factor 3/fork head or "winged helix" proteins: a family of transcription factors of diverse biologic function. *Proc. Natl Acad. Sci. USA* **90**, 10421–10423 (1993).
- Li, C. & Tucker, P. W. DNA-binding properties and secondary structural model of the hepatocyte nuclear factor 3/fork head domain. *Proc. Natl Acad. Sci. USA* **90**, 11583–11587 (1993).
- Kaufmann, E. & Knöchel, W. Five years on the wings of fork head. *Mech. Dev.* **57**, 3–20 (1996).
- Kaestner, K. H., Knöchel, W. & Martinez, D. E. Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev.* **14**, 142–146 (2000).
- Shu, W., Yang, H., Zhang, L., Lu, M. M. & Morrissey, E. E. Characterization of a new subfamily of winged-helix/forkhead (fox) genes that are expressed in the lung and act as transcriptional repressors. *J. Biol. Chem.* **276**, 27488–27497 (2001).
- Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**, 412–420 (1993).
- Nishimura, D. Y. *et al.* The forkhead transcription factor gene FKHL7 is responsible for glaucoma phenotypes which map to 6p25. *Nature Genet.* **19**, 140–147 (1998).
- Mears, A. J. *et al.* Mutations of the forkhead/winged-helix gene, FKHL7, in patients with Axenfeld-Rieger anomaly. *Am. J. Hum. Genet.* **63**, 1316–1328 (1998).
- Clifton-Bligh, R. J. *et al.* Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia. *Nature Genet.* **19**, 399–401 (1998).
- Fang, J. *et al.* Mutations in FOXC2 (MFH-1), a forkhead family transcription factor, are responsible for the hereditary lymphedema-distichiasis syndrome. *Am. J. Hum. Genet.* **67**, 1382–1388 (2000).
- Crisponi, L. *et al.* The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. *Nature Genet.* **27**, 159–166 (2001).
- Semina, E. V., Brownell, I., Mintz-Hittner, H. A., Murray, J. C. & Jamrich, M. Mutations in the human forkhead transcription factor FOXE3 associated with anterior segment ocular dysgenesis and cataracts. *Hum. Mol. Genet.* **10**, 231–236 (2001).
- Widm, R. S. *et al.* X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse *scuffy*. *Nature Genet.* **27**, 18–20 (2001).



encoded by any *Yersinia* plasmid, but horizontal mobility has apparently occurred: a plasmid closely related to pFra (but lacking *ymt* and the *caf* operon) exists in the exclusively human pathogen *Salmonella enterica* serovar Typhi⁹.

Many regions within the *Y. pestis* chromosome showed some of the characteristics of islands acquired through lateral transfer (see Supplementary Information). Among these were several genes that seem to have come from other insect pathogens. Only two features of *Y. pestis* have so far been shown to be essential for aspects of its life cycle in the flea: the plasmid-encoded murine toxin Ymt is essential for flea colonization¹⁰, and the chromosomal

hms locus is required for blockage of the flea midgut by *Y. pestis* to maximize its transmission¹. This second locus is also present in *Y. pseudotuberculosis*¹¹.

Sequences related to the parasitism of insects include genes encoding homologues of the insecticidal toxin complexes (Tcs) from *Photobacterium luminescens*, *Serratia entomophila* and *Xenorhabdus nematophilus*¹². These toxins are complexes of the products of three different gene families: *tcaA/tcaB/tcdA*, *tcaC/tcdB* and *tccC*. In *Y. pestis*, three adjacent genes encoding homologues of *P. luminescens* TcaA, TcaB and TcaC were identified (YPO3681, 35% identity; YPO3679, 42% identity; and YPO3678,

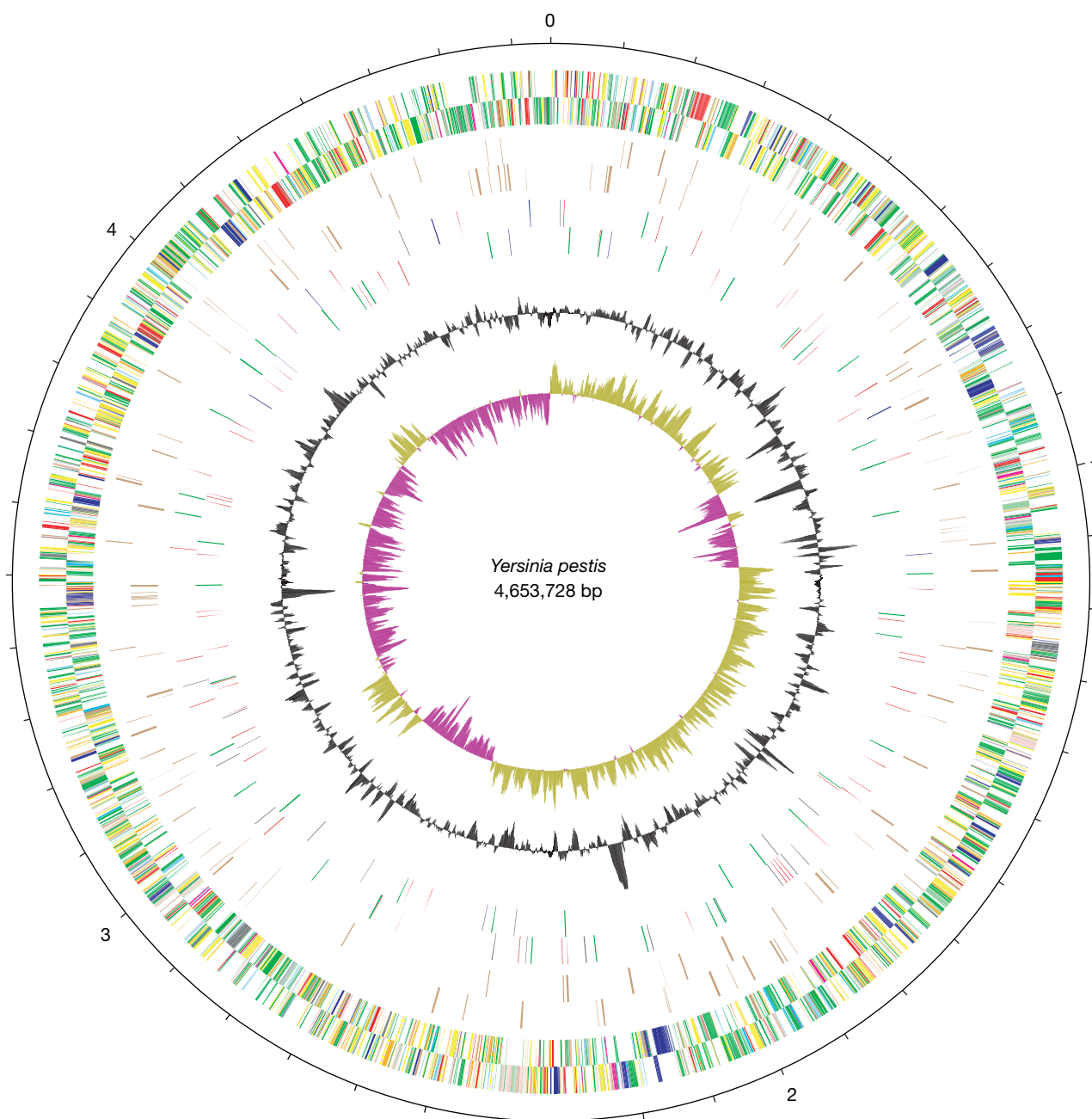


Figure 1 Circular representation of the *Y. pestis* genome. The outer scale is marked in megabases. Circles 1 and 2 (from the outside in), all genes colour coded by function, forward and reverse strand; circles 3 and 4, pseudogenes; circles 5 and 6, insertion sequence elements (blue, IS1661; black, IS285; red, IS1541; green, IS100); circle 7, G + C content (higher values outward); circle 8, GC bias ((G – C/G + C), khaki indicates values >1, purple <1). Colour coding for genes: dark blue, pathogenicity or adaptation;

black, energy metabolism; red, information transfer; dark green, surface associated; cyan, degradation of large molecules; magenta, degradation of small molecules; yellow, central or intermediary metabolism; pale blue, regulators; orange, conserved hypothetical; brown, pseudogenes; pink, phage and insertion sequence elements; pale green, unknown; grey, miscellaneous.

Table 1 General features of the *Yersinia pestis* genome

	Chromosome	pPst/pPCP1	pYV1/pCD1	pFra/pMT1
Estimated copy number*		186	4.3	1.8
Total size	4,653,728 bp	9,612 bp	70,305 bp	96,210 bp
G + C content	47.64%	45.27%	44.84%	50.23%
Coding sequences	4,012	9	97	103
... of which pseudogenes	149	0	8	3
Coding density	83.8%	57.2%	81.4%	86.8%
Average gene length	998 bp	611 bp	643 bp	835 bp
Ribosomal RNAs	6 × (16S–23S–5S)			
Transfer RNAs	70			
Other stable RNAs	6			

*The copy number of the plasmids was estimated on the number of reads per kilobase compared with the chromosome, and is likely to be affected by sampling bias in the shotgun.

49% identity) separated from nearby homologues of TccC (YPO3673, 52% identity; and YPO3674, 54% identity) by phage-related genes. The *tcaA* gene was intact, but *tcaB* contained a frameshift mutation and *tcaC* an internal deletion. The disruption of these genes might be necessary for the lifestyle of *Y. pestis*, which persists in the flea gut for relatively long periods. Two isolated genes encoding homologues of TccC were also present in the genome (YPO2312, 60% identity; and YPO2380, 74% identity). Strikingly, a protein showing weak, but significant, similarity only to the enhancins encoded uniquely by baculoviral pathogens of insects was identified (YPO0339, 25% identity) in a region of low G + C content, flanked by a transfer RNA gene and transposase fragments, suggesting horizontal acquisition. During baculoviral pathogenesis, the proteolytic activity of enhancin damages the peritrophic membrane, which normally provides a physical barrier against microbial pathogens in the insect midgut¹³ and the product of YPO0339 may be important in colonization of the flea by *Y. pestis*. Interestingly,

PCR shows (R.W.T., data not shown) that the *tca* insecticidal toxin genes are also present in *Y. pseudotuberculosis* IP 32953 (virulent serotype I strain). This suggests an association between *Y. pseudotuberculosis* and insects, or insect pathogens in the environment, before the emergence of *Y. pestis*.

The type III secretion system located on the *Yersinia* virulence plasmid (pYV/pCD1) is present in all three human pathogenic species of *Yersinia*. This system allows translocation of a range of effector proteins (Yops) that downregulate the responses of host phagocytic cells to infection. Human pathogenic *Yersinia* possessing mutations that disrupt this system are severely attenuated¹⁴. A second chromosomally encoded type III secretion system, similar in gene content and order to the SPI-2 type III system of *Salmonella typhimurium*¹⁵, is also present in *Y. pestis* (Fig. 2). This secretion system is distinct from the chromosomally borne type III system in *Y. enterocolitica*¹⁶. We were unable to identify the effectors for this system, as these are generally divergent.

The genome sequence of *Y. pestis* CO92 also revealed the presence of a range of genes predicted to encode previously unknown surface antigens, which might have a role in the virulence of the bacterium. Two fimbrial-type surface structures were known in *Y. pestis*: the locus encoding pH6 antigen (*psa*) is chromosomally borne and is also found in *Y. pseudotuberculosis*, and the *caf* locus, encoding the F1 antigen, is found only in *Y. pestis* on plasmid pFra. We also identified eight systems with a similar organization to the *psa* and *caf* operons, each of which provides potential mechanisms for fimbrial or adhesin production (Fig. 2). In five cases the operons are flanked by genes encoding transposases or integrases, again implying horizontal acquisition. Such high redundancy of fimbria-related genes is also found in other bacterial pathogens, including

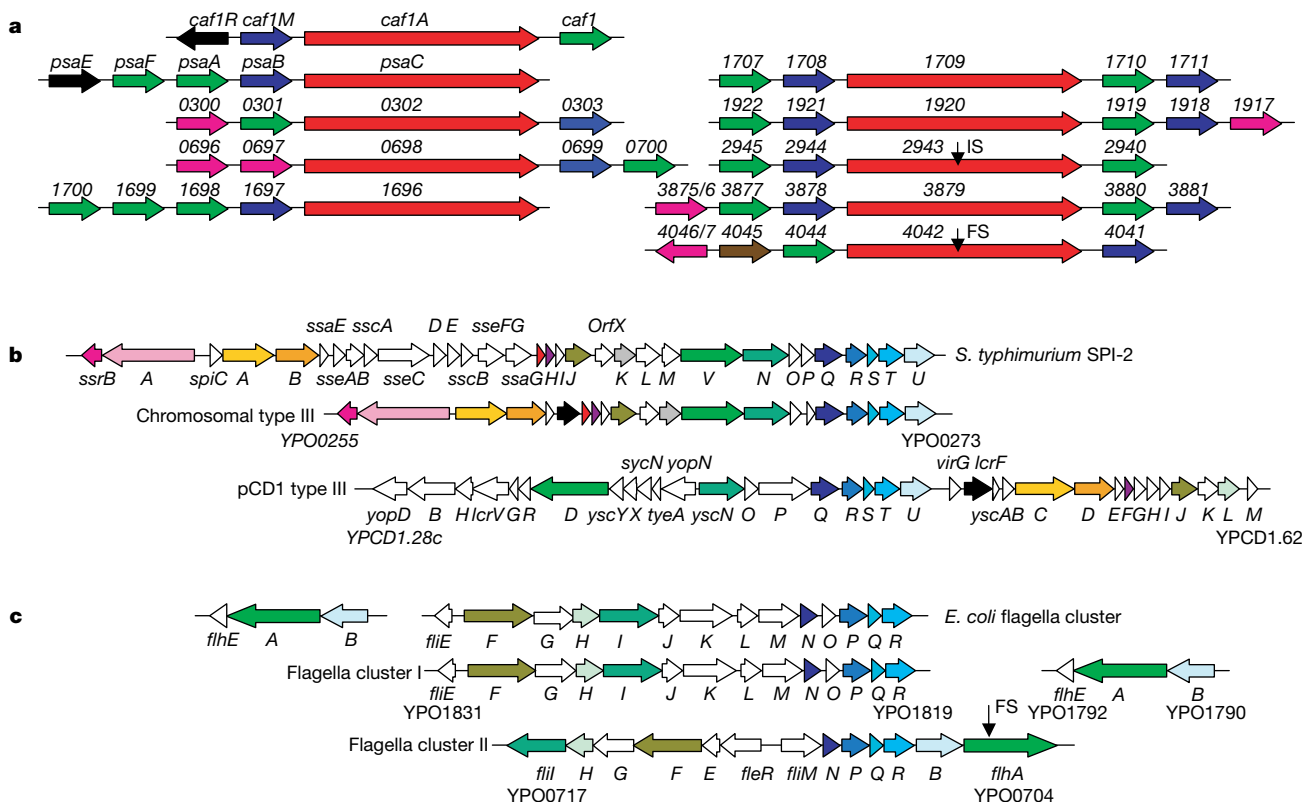


Figure 2 Protein secretion systems in *Y. pestis*. **a**, Chaperone–usher systems. The plasmid pMT1-borne *caf* system and nine chromosomal systems are shown. Blue, chaperones; red, usher proteins; green, putative target proteins; black, regulatory proteins; pink, transposases and integrases; brown, other genes. IS, insertion sequence element; FS, frameshift. **b**, Type III secretion systems. The chromosomal

and plasmid type III systems are shown, with *S. typhimurium* SPI-2 as a comparator. **c**, Flagella operons. The parts of the flagella clusters with similarity to type III systems are shown, with the *E. coli* flagella cluster as a comparator. The genes are arbitrarily colour coded in **b** and **c** to show related genes.

Escherichia coli and *Salmonella enterica* serovar Typhi¹⁷. A large arsenal of independent gene clusters encoding different fimbriae and adhesins might be beneficial in evading host immune response, or may allow multiple interactions with several different hosts during its complex life cycle.

Although there is ample evidence of horizontal gene acquisition, many of these sequences will have been acquired before *Y. pestis* evolved from *Y. pseudotuberculosis*. Preliminary comparisons with the unfinished *Y. pseudotuberculosis* sequence from the Lawrence Livermore National Laboratories (<http://bbrrp.llnl.gov/bbrp/html/microbe.html>) indicate that around 40% of the islands identified may be partially or completely absent from *Y. pseudotuberculosis*, although this must be treated with caution, given the incomplete nature of the data. Phylogenetic analysis to accurately date the acquisition of these genetic islands will have to await the completion of the genome sequences of *Y. pseudotuberculosis* and *Y. enterocolitica*.

Horizontal gene acquisition in *Y. pestis* has been balanced by gene loss. In total the genome sequence contains 149 pseudogenes, of which 51 are a consequence of disruption by insertion sequence elements. The total number of insertion sequences exceeds that described in most other bacterial genomes, comprising 3.7% of the genome. At least four different insertion sequences (IS) were found on the chromosome: 66 complete or partial copies of IS1541, 44 of IS100, 21 of IS285 and 9 of IS1661. Overall numbers of insertion sequence copies are around tenfold higher than in *Y. pseudotuberculosis* (IS1541, 7–13 copies¹⁸; IS100, 0–6 copies¹⁹). Fifty-eight pseudogenes were due to frameshift mutations (21 at homopolymeric tracts), 32 due to deletions, and the remainder due to in-frame stop codons. Mutations in genes associated with pathogenicity were over-represented (see Supplementary Information). Plasmid pCD1, which is shared with the enteropathogens *Y. enterocolitica* and *Y. pseudotuberculosis*, contained mutations in virulence-related genes *ylpA* and *yadA*, and in five other genes, but of the unique *Y. pestis* plasmids, pMT1 has only two pseudogenes and pPst has none.

The change in lifestyle of *Y. pestis* compared with the ancestral *Y. pseudotuberculosis* strain would be expected to result in the loss of genes required for enteropathogenicity. Enteropathogens specifically adhere to surfaces of the gut and may invade cells lining it. Proteins important for this process in *Y. pseudotuberculosis* include YadA and Invasin, both of which are represented by pseudogenes in *Y. pestis* (see Supplementary Information). Several of the other pseudogenes reported here could encode adhesin molecules, suggesting that some of these might also have been required for enteropathogenicity. The pseudogene YPO1562 shows 34% amino-acid identity with *E. coli* intimin, which is carried on a pathogenicity island termed the locus of enterocyte effacement. *Yersinia pseudotuberculosis* also contains an intact gene sharing 61% amino-acid identity with a further *E. coli* virulence factor, cytotoxic necrotizing factor 1 (CNF1); this is also represented by a pseudogene (YPO1449) in *Y. pestis*. CNF1 acts on Rho GTPases to affect cytoskeletal rearrangement²⁰, which may be required for epithelial invasion.

Thirty-eight (26%) of the pseudogenes would have encoded or synthesized surface-expressed antigens or exported proteins. Several mucosal pathogens have been shown to switch surface-expressed antigens on or off *in vitro* and *in vivo* using slipped-strand mispairing of repeat sequences during replication²¹. A similar process has been demonstrated in *Y. pestis*. The organism is characteristically urease negative but activity can be restored *in vitro* by the spontaneous deletion of a single base pair in a homopolymeric tract in *ureD*²². This type of reversible mutation would reduce the metabolic burden of producing proteins unnecessary to *Y. pestis* in its new flea/mammal life cycle yet still allow the potential to express these should a subsequent need arise.

Some typical virulence properties of enteropathogens are asso-

ciated with systems that in *Y. pestis* are subject to multiple mutations which would make reversion unlikely. Motility is required for efficient invasion of host cells by *Y. enterocolitica*²³, apparently by promoting bacterial contact with the host cell; strains of *Y. pestis* are uniformly nonmotile, and analysis of the flagellar and chemotaxis gene clusters reveal six mutations. However, there are two separate and distinct flagellar gene clusters (Fig. 2), one of which contains no obvious mutations in flagellar biosynthetic or structural genes, suggesting that some form of motility may be possible under as-yet-undiscovered conditions.

Five pseudogenes were lipopolysaccharide (LPS) biosynthesis genes. The O side chain of LPS is an important factor in the virulence of a range of pathogens, including *Y. enterocolitica*²⁴. The O side chain mediates resistance to complement-mediated and phagocyte killing. It might also have a role in survival in the gut by protecting the bacterium from cationic peptides, such as those produced by Paneth cells in the human small intestine²⁵. *Yersinia pestis* produces a rough LPS, lacking an O antigen as a consequence of these mutations within the biosynthesis cluster of the O antigen³, but the nature of the selective advantage this may confer to *Y. pestis* is unexplained. The surface adhesin Ail (YPO2905) is also involved in serum resistance in *Y. enterocolitica*²⁶. An IS285 insertion was reported in the *ail* gene of a laboratory-adapted strain of *Y. pestis* used as a vaccine²⁷, and *ail* is generally assumed to be inactivated in *Y. pestis*⁸. However, in strain CO92 it is intact. Three other intact genes encoding different Ail-like proteins were identified (YPO1850, YPO2190, YPO2506); *Y. pestis* expresses an unidentified plasmid-independent adhesin⁸ that could therefore be Ail or a paralogue. Rough mutants of *Y. enterocolitica* show an enhanced ability to invade cultured cells, a suggested consequence of the enhanced accessibility of surface Ail²⁸. However, a recent mutagenesis study in *Y. pseudotuberculosis* found that O-antigen mutants were unable to invade epithelial cells²⁹ and were attenuated when infecting mice both orally and systemically, suggesting that as *Y. pestis* is invasive⁸ and virulent, it may have alternative surface structures that complement the loss of the O antigen.

Mutations in energy metabolism and central and intermediary metabolism are comparatively rare (see Supplementary Information). A mutation in the methionine biosynthetic pathway, and in *pheA*, may account for known requirements for growth *in vitro* of one or more of the amino acids cysteine, methionine and phenylalanine¹. In contrast, many of the mutations were in uptake and transport systems, indicating that, compared with life as a gastrointestinal pathogen, fewer or different types of nutrients are available to *Y. pestis* within the flea or mammal. Three isolated genes possibly involved in iron uptake and a gene essential for aerobactin production (*iucA*, YPO0989) are inactivated in *Y. pestis* CO92. However, adding to previously described iron-uptake systems, we identified three apparently functional siderophore biosynthesis operons, five siderophore uptake systems, a non-siderophore iron-uptake system and a second haem-acquisition system.

Several mechanisms account for the accumulation of pseudogenes in *Y. pestis*, including expansion of insertion sequence elements, deletion, point mutation and slippage in homopolymeric tracts. Together, these have resulted in the loss of genes that were either not required in its new niche, or may have decreased fitness for a new lifestyle as a systemic pathogen of mammals and an insect pathogen. Comparison of the pseudogenes (excluding those in insertion sequences and phages) with the unfinished *Y. pseudotuberculosis* sequence (see above) shows that, of the >60% for which information is available, >95% of the *Y. pseudotuberculosis* orthologues do not carry the inactivating mutation. This, combined with the published evidence of intact genes in *Y. pseudotuberculosis* detailed above, indicates that most of the mutations in *Y. pestis* are recent. They may have arisen in a gradual process since the colonization of this new niche, or *Y. pestis* could have acquired these pseudogenes as a consequence of an evolutionary bottleneck that

may have accompanied the colonization process.

The genome sequence of *Y. pestis* reveals a pathogen that has undergone considerable genetic flux, with evidence of selective genome expansion by lateral gene transfer of plasmid and chromosomal genes, and subsequent initial stages of genome size reduction. We believe that these features correlate with a change in pathogenic niche, and therefore this genome sequence provides a unique insight into the genetic events that are associated with the emergence of a new pathogenic species. The newly emerged pathogen is highly virulent for humans, causing pandemics of systemic, and often fatal, disease, contrasting with the ancestral species that evolved to cause non-fatal enteritis in similar hosts. □

Methods

A single colony of *Y. pestis* strain CO92 was picked from Congo Red agar and grown overnight in BAB broth with shaking at 37 °C. Cells were collected and total DNA (10 mg) was isolated using proteinase K treatment followed by phenol extraction. The DNA was fragmented by sonication, and several libraries were generated in pUC18 using size fractions ranging from 1.0 to 2.5 kb. The whole genome sequence was obtained from 94,881 end sequences (giving 9.6× coverage) derived from these libraries using dye terminator chemistry on ABI377 automated sequencers. End sequences from larger insert plasmid (pSP64; 3.1× clone coverage, 9–11 kb insert size) and lambda (lambda-FIX-IL; 4.0× clone coverage, 20–22 kb insert size) libraries were used as a scaffold. The sequence was assembled, finished and annotated as described previously³⁰, using the program Artemis (<http://www.sanger.ac.uk/Software/Artemis>) to collate data and facilitate annotation. The genome sequences of *Y. pestis* and *E. coli* were compared pairwise using the Artemis Comparison Tool (ACT) (<http://www.sanger.ac.uk/Software/ACT>). Pseudogenes had one or more mutations that would ablate expression; each of the inactivating mutations was subsequently checked against the original sequencing data.

Received 8 May; accepted 16 August 2001.

- Perry, R. D. & Fetherston, J. D. *Yersinia pestis*—etiologic agent of plague. *Clin. Microbiol. Rev.* **10**, 35–66 (1997).
- Galimand, M. *et al.* Multidrug resistance in *Yersinia pestis* mediated by a transferable plasmid. *N. Engl. J. Med.* **337**, 677–680 (1997).
- Skurumik, M., Peippo, A. & Ervela, E. Characterization of the O-antigen gene clusters of *Yersinia pseudotuberculosis* and the cryptic O-antigen gene cluster of *Yersinia pestis* shows that the plague bacillus is most closely related to and has evolved from *Y. pseudotuberculosis* serotype O:1b. *Mol. Microbiol.* **37**, 316–330 (2000).
- Achtman, M. *et al.* *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc. Natl Acad. Sci. USA* **96**, 14043–14048 (1999).
- Doll, J. M. *et al.* Cat-transmitted fatal pneumonic plague in a person who traveled from Colorado to Arizona. *Am. J. Trop. Med. Hyg.* **51**, 109–114 (1994).
- Lobry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13**, 660–665 (1996).
- Guiyoule, A. *et al.* Recent emergence of new variants of *Yersinia pestis* in Madagascar. *J. Clin. Microbiol.* **35**, 2826–2833 (1997).
- Cowan, C., Jones, H. A., Kaya, Y. H., Perry, R. D. & Straley, S. C. Invasion of epithelial cells by *Yersinia pestis*: evidence for a *Y. pestis*-specific invasin. *Infect. Immun.* **68**, 4523–4530 (2000).
- Prentice, M. B. *et al.* *Yersinia pestis* pFra shows biovar-specific differences and recent common ancestry with a *Salmonella enterica* serovar Typhi plasmid. *J. Bacteriol.* **183**, 2586–2594 (2001).
- Hinnebusch, J. *et al.* Murine toxin of *Yersinia pestis* shows phospholipase D activity but is not required for virulence in mice. *Int. J. Med. Microbiol.* **290**, 483–487 (2000).
- Buchrieser, C. *et al.* The 102-kilobase *pgm* locus of *Yersinia pestis*: sequence analysis and comparison of selected regions among different *Yersinia pestis* and *Yersinia pseudotuberculosis* strains. *Infect. Immun.* **67**, 4851–4861 (1999).
- Waterfield, N. R., Bowen, D. J., Fetherston, J. D., Perry, R. D. & French-Constant, R. H. The *tc* genes of *Photobacterium*: a growing family. *Trends Microbiol.* **9**, 185–191 (2001).
- Wang, P. & Granados, R. R. An intestinal mucin is the target substrate for a baculovirus enhancer. *Proc. Natl Acad. Sci. USA* **94**, 6977–6982 (1997).
- Cornelis, G. R. *et al.* The virulence plasmid of *Yersinia*, an antihist genome. *Microbiol. Mol. Biol. Rev.* **62**, 1315–1352 (1998).
- Shea, J. E., Hensel, M., Gleeson, C. & Holden, D. W. Identification of a virulence locus encoding a second type III secretion system in *Salmonella typhimurium*. *Proc. Natl Acad. Sci. USA* **93**, 2593–2597 (1996).
- Haller, J. C., Carlson, S., Pederson, K. J. & Pierson, D. E. A chromosomally encoded type III secretion pathway in *Yersinia enterocolitica* is important in virulence. *Mol. Microbiol.* **36**, 1436–1446 (2000).
- Townsend, S. M. *et al.* *Salmonella enterica* serovar Typhi possesses a unique repertoire of fimbrial gene sequences. *Infect. Immun.* **69**, 2894–2901 (2001).
- Odaert, M., Berche, P. & Simonet, M. Molecular typing of *Yersinia pseudotuberculosis* by using an IS200-like element. *J. Clin. Microbiol.* **34**, 2231–2235 (1996).
- McDonough, K. A. & Hare, J. M. Homology with a repeated *Yersinia pestis* DNA sequence IS100 correlates with pesticin sensitivity in *Yersinia pseudotuberculosis*. *J. Bacteriol.* **179**, 2081–2085 (1997).
- Lerm, M., Schmidt, G. & Aktories, K. Bacterial protein toxins targeting rho GTPases. *FEMS Microbiol. Lett.* **188**, 1–6 (2000).
- Henderson, I. R., Owen, P. & Nataro, J. P. Molecular switches—the ON and OFF of bacterial phase variation. *Mol. Microbiol.* **33**, 919–932 (1999).
- Sebbane, F., Devalckenaere, A., Foulon, J., Carniel, E. & Simonet, M. Silencing and reactivation of urease in *Yersinia pestis* is determined by one G residue at a specific position in the *ureD* gene. *Infect. Immun.* **69**, 170–176 (2001).
- Young, G. M., Badger, J. L. & Miller, V. L. Motility is required to initiate host cell invasion by *Yersinia*

enterocolitica. *Infect. Immun.* **68**, 4323–4326 (2000).

- Darwin, A. J. & Miller, V. L. Identification of *Yersinia enterocolitica* genes affecting survival in an animal host using signature-tagged transposon mutagenesis. *Mol. Microbiol.* **32**, 51–62 (1999).
- Nesper, J. *et al.* Characterization of *Vibrio cholerae* O1 El tor *galU* and *galE* mutants: influence on lipopolysaccharide structure, colonization, and biofilm formation. *Infect. Immun.* **69**, 435–445 (2001).
- Wachtel, M. R. & Miller, V. L. *In vitro* and *in vivo* characterization of an *ail* mutant of *Yersinia enterocolitica*. *Infect. Immun.* **63**, 2541–2548 (1995).
- Torosian, S. & Zsigray, R. In Abstracts of the 96th General Meeting of the American Society of Microbiology 1996 Abstract B-13, 191 (American Society for Microbiology, Washington DC, 1996).
- Pierson, D. E. Mutations affecting lipopolysaccharide enhance *ail*-mediated entry of *Yersinia enterocolitica* into mammalian cells. *J. Bacteriol.* **176**, 4043–4051 (1994).
- Mecas, J., Bilis, I. & Falkow, S. Identification of attenuated *Yersinia pseudotuberculosis* strains and characterization of an orogastric infection in BALB/c mice on day 5 postinfection by signature-tagged mutagenesis. *Infect. Immun.* **69**, 2779–2787 (2001).
- Parkhill, J. *et al.* Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404**, 502–506 (2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We acknowledge the support of the Sanger Centre core sequencing, subcloning and informatics groups. This work was funded by the Wellcome Trust through its Beowulf Genomics initiative. We are grateful to E. Garcia of the *Y. pseudotuberculosis* sequencing group at the Lawrence Livermore National Laboratories for giving permission to use sequence data before they are published. B.W.W. and A.V.K. are supported by the Wellcome Trust.

Correspondence and requests for materials should be addressed to J.P. (e-mail: parkhill@sanger.ac.uk) or B.W.W. (e-mail: brendan.wren@lshtm.ac.uk). The sequences have been submitted to EMBL under the accession numbers AL590842 (chromosome), AL109969 (pPCP1), AL117189 (pCD1) and AL117211 (pMT1).

An endogenous cannabinoid (2-AG) is neuroprotective after brain injury

David Panikashvili^{*†}, Constantina Simeonidou^{*}, Shimon Ben-Shabat[†], Lumir Hanuš[†], Aviva Breuer[†], Raphael Mechoulam[†] & Esther Shohami^{*}

^{*} Department of Pharmacology, and [†] Department of Medicinal Chemistry and Natural Products, Medical Faculty, Hebrew University, Jerusalem 91120, Israel

Traumatic brain injury triggers the accumulation of harmful mediators that may lead to secondary damage^{1,2}. Protective mechanisms to attenuate damage are also set in motion². 2-Arachidonoyl glycerol (2-AG) is an endogenous cannabinoid, identified both in the periphery³ and in the brain⁴, but its physiological roles have been only partially clarified^{5–7}. Here we show that, after injury to the mouse brain, 2-AG may have a neuroprotective role in which the cannabinoid system is involved. After closed head injury (CHI) in mice, the level of endogenous 2-AG was significantly elevated. We administered synthetic 2-AG to mice after CHI and found significant reduction of brain oedema, better clinical recovery, reduced infarct volume and reduced hippocampal cell death compared with controls. When 2-AG was administered together with additional inactive 2-acyl-glycerols that are normally present in the brain, functional recovery was significantly enhanced. The beneficial effect of 2-AG was dose-dependently attenuated by SR-141761A, an antagonist of the CB₁ cannabinoid receptor.

Traumatic brain injury is a major cause of mortality and morbidity, yet there is no effective drug to treat brain-injured patients. Understanding post-trauma events and developing neuroprotective agents are therefore important⁸. Cannabinoids act through two distinct receptors, one of which (CB₁) is found mainly in the brain. We have previously reported that 2-AG suppresses formation of reactive oxygen species (ROS) and tumour necrosis factor- α (TNF- α) by murine macrophages

in vitro after stimulation with lipopolysaccharide (LPS), and have noted lower levels of TNF- α in the serum of LPS-treated mice after administration of 2-AG⁹. Both classes of mediators, ROS¹⁰ and TNF- α ¹¹, are major contributors to the pathophysiology of brain injury.

On the basis of these observations, we assumed that 2-AG could have a neuroprotective role in the post-traumatic brain. We therefore investigated the dynamic changes in brain 2-AG levels after traumatic brain injury; the possibility that exogenous 2-AG may attenuate brain damage after CHI; and the involvement of the CB₁ receptor in neuroprotection.

Mice were subjected to CHI using a weight-drop device, or to sham surgery, as described elsewhere¹². They were decapitated at various time intervals (15 min, 1, 4, 8 or 24 h) and the brains were removed within 1 min and frozen in liquid nitrogen. The rapid freezing minimized the post-mortem production of 2-AG. This was confirmed by the similar levels of 2-AG obtained in brains extracted at 1, 3 or 5 min after decapitation (data not shown). The levels of brain 2-AG were determined in the traumatized hemispheres (Fig. 1) and compared with those in controls. One hour after CHI the levels of 2-AG were already significantly higher than in sham, non-injured mice, peaking at a tenfold increase after 4 h, and declining thereafter. Even after 24 h, the levels of 2-AG were still higher (sixfold) than in controls. In the contralateral hemisphere, the changes in 2-AG were minor, and not significant (data not shown).

To test the role of 2-AG in post-traumatic pathophysiology, we synthesized 2-AG as previously described³ and intravenously administered doses of 0.1, 5 and 10 mg kg⁻¹ 15 min after CHI. One of the early manifestations of brain injury is oedema formation (accumulation of water in the tissue), leading to increased intracranial pressure. We therefore examined the effect of 2-AG on oedema formation by measuring water content 24 h after trauma, the time of maximal oedema¹². At all doses tested, 2-AG significantly reduced water accumulation by about 50% (Fig. 2a).

The clinical status of the mice was evaluated 1 and 24 h after CHI using a scoring system testing motor and behavioural functions¹³ (neurological severity score, NSS: 0, healthy, to 10, fatally injured). The initial NSS (at 1 h) reflects the severity of injury, and Δ NSS, the difference between the score at 1 and 24 h, reflects recovery. Whereas the NSS (1 h) was similar in the treated and control groups (8.25 ± 0.36 and 7.63 ± 0.46 , respectively), the mice treated with 2-AG recovered faster within the first 24 h. All doses of 2-AG showed a trend towards better recovery (greater Δ NSS), but it reached

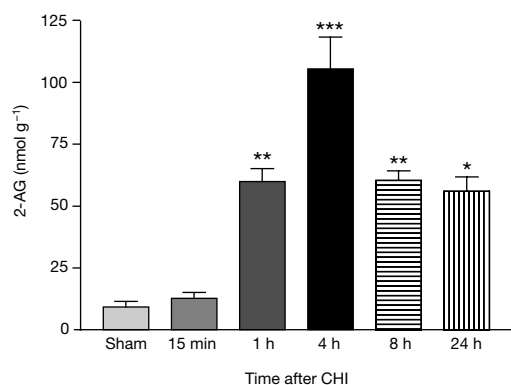


Figure 1 Temporal profile of 2-AG changes after CHI. CHI was induced over the left cerebral hemisphere. At the designated times the mice were killed and 2-AG was extracted from the contused hemisphere, separated on TLC and analysed by GC-MS. 2-AG levels were already significantly elevated 1 h after CHI, peaking at 4 h and sustained until 24 h. Controls are sham, non-injured mice. An analysis of variance (ANOVA) ($F = 36.01$, $P < 0.001$) with Tukey's *post hoc* test were performed: single asterisk, $P < 0.05$ versus control; double asterisk, $P < 0.01$ versus control; triple asterisk, $P < 0.001$ versus control. Data are mean $\pm$ s.e.m.

significance only at 5 mg kg⁻¹ ($P < 0.01$, versus control; Fig. 2b). The effect of 2-AG was not long lasting. We injected 2-AG 1 h after CHI, and evaluated NSS 1, 4 and 7 days later, but found significant protection on the first day only, and not on subsequent days (Fig. 2c). Brain oedema and neurological deficits are not necessarily interdependent, because they represent different manifestations of brain damage. Indeed, we found that 2-AG is less potent against the motor deficits than in reducing oedema, as evidenced from the different effective doses needed to achieve protection.

One of the manifestations of brain injury is neuronal cell death, which is grossly evidenced by an infarct around the site of injury. Mice were treated with either 2-AG (5 mg kg⁻¹, $n = 7$) or vehicle ($n = 12$), decapitated 24 h later and their brains were examined for changes in infarct volume¹⁴. Infarct volume was significantly reduced in 2-AG-treated mice compared with that in vehicle-treated mice ($5.6 \pm 1.2\%$ and $10.2 \pm 1.1\%$, respectively, of total brain volume; $P = 0.021$). No infarct was observed in sham-operated mice (data not shown). In addition, we examined the effect of 2-AG on hippocampal cell death, which occurs specifically in the CA3 subfield¹². Mice were treated with vehicle or with 5 mg kg⁻¹ of 2-AG

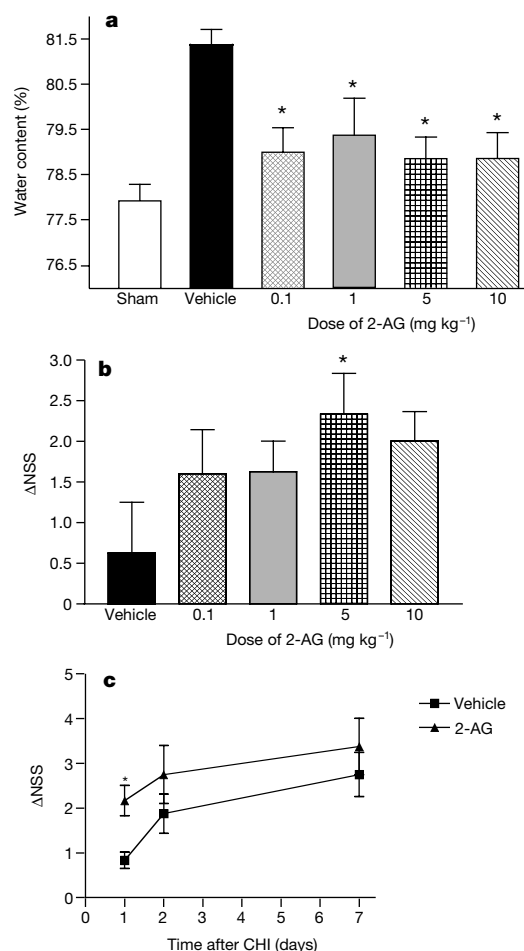


Figure 2 Protective effects of 2-AG after CHI. 2-AG was administered after CHI, and its protection against oedema formation (a) and effect on clinical recovery (b, c) are shown. a, 2-AG, at a dose range of 0.1–10 mg kg⁻¹, was injected 15 min after CHI, and tissue water content was evaluated 24 h later in the contused brain hemisphere. At all doses, 2-AG significantly reduced oedema. Controls are sham, non-injured mice and CHI-traumatized mice injected with vehicle. An ANOVA ($F = 5$, 185, $P < 0.001$) with Tukey's *post hoc* test were performed: asterisk, $P < 0.05$ versus control vehicle-injected CHI mice. b, Δ NSS reflects clinical recovery: a higher value indicates greater recovery. c, 2-AG (5 mg kg⁻¹) was injected 1 h after CHI and Δ NSS was assessed at 24 h, 48 h and 7 days. Data are mean $\pm$ s.e.m.

1 h after CHI, and killed 7 days later by perfusion fixation. Serial brain sections were stained with haematoxylin and eosin and Nissl. The ipsilateral hippocampus of vehicle- and 2-AG-treated mice showed ischaemic and dystrophic neurons, with large or small patchy acellular regions, in the CA3 subfield (Fig. 3). No neuronal pathology could be detected in the contralateral CA3 region (Fig. 3a, d). Neuronal loss was apparent in the injured side compared with the contralateral side of both vehicle- and 2-AG-treated mice (Fig. 4a); however, this loss was significantly more obvious in the vehicle-treated animals. The number of neurons counted in the CA3 area in the injured hemisphere, expressed as a percentage of those counted in the CA3 area in the intact hemisphere, was significantly lower in the vehicle-treated mice than in the 2-AG-treated mice (Fig. 4b).

2-AG is accompanied in mouse brains by an 'entourage' of congeners, such as 2-palmitoyl-glycerol and 2-linoleoyl-glycerol, which alone have no activity at the cannabinoid receptors. However, these glycerol esters facilitate the binding of 2-AG to the CB₁ and CB₂ receptors, probably through increasing the availability of 2-AG at the receptors by inhibiting its hydrolysis and uptake¹⁵. To test the effect of the entourage after CHI, mice were injected with a combination of 2-AG (1 mg kg⁻¹), 2-palmitoyl-glycerol (5 mg kg⁻¹) and 2-linoleoyl-glycerol (10 mg kg⁻¹) 1 h after CHI. Although no effect was observed when these compounds were injected separately, the combination of the three acyl-glycerol esters led to a significant improvement in functional recovery at 24 h (Fig. 5a). When the combination (2-AG plus entourage) was administered again 24 and 48 h after CHI, the improved clinical recovery was sustained up to 7 days after the trauma (Fig. 5b).

To determine whether the protective effect of the exogenous 2-AG is mediated at least in part by the CB₁ cannabinoid receptor, we injected SR-141716A, an antagonist specific to the CB₁ receptor subtype, after CHI, along with 2-AG. Oedema was determined 24 h later. Five groups of mice were subjected to CHI and treated with

either vehicle (control), 2-AG (1 mg kg⁻¹) alone or with 3, 10 and 20 mg kg⁻¹ of SR-141716A. The protective effect of 2-AG was dose-dependently attenuated by the antagonist (Fig. 6). The water content in mice treated with 2-AG and 20 mg kg⁻¹ of antagonist did not differ from that in the CHI control mice. The doses of the antagonist used were relatively high; however, the effective dose could depend upon the species or strain of animal used in the study. High doses (≈20 mg kg⁻¹) of the antagonist have previously been used^{16,17}. It should be noted, however, that the antagonist alone did not worsen the outcome of traumatized mice not treated with 2-AG (data not shown).

These results are, to our knowledge, the first to demonstrate temporal and local changes in brain levels of the endogenous cannabinoid 2-AG after traumatic brain injury and to record its neuroprotective effects. We have previously reported, using the same model of CHI, that the arachidonic acid metabolic cascade is activated after trauma¹⁸, and that there is massive accumulation of calcium at the site of injury¹⁹. Both events may be related to 2-AG release. It seems that 2-AG can be synthesized and released from neuronal and non-neuronal cells on demand by a non-synaptic release mechanism⁵. Endocannabinoids are taken up by cells using a specific transporter²⁰ or, in particular for 2-AG, also by passive diffusion⁵ and are hydrolysed by fatty acid amide hydrolase (FAAH), which acts on both anandamide and 2-AG^{5,21}.

Several cannabinoids that bind to CB₁ have been assayed for neuroprotection (mostly using *in vitro* models). 2-AG has been shown to protect cerebral neurons of rats from ischaemia *in vitro*²², and cannabinoid receptor agonists inhibit glutamatergic synaptic transmission in hippocampal cultures *in vitro*²³. The synthetic cannabinoid WIN 55212-2 decreases hippocampal neuronal loss after transient global cerebral ischaemia induced by occlusion of the middle cerebral artery in rats²⁴. Indeed, our present findings show significant protection of hippocampal cells after treatment with 2-AG. Tetrahydrocannabinol was also shown to protect against

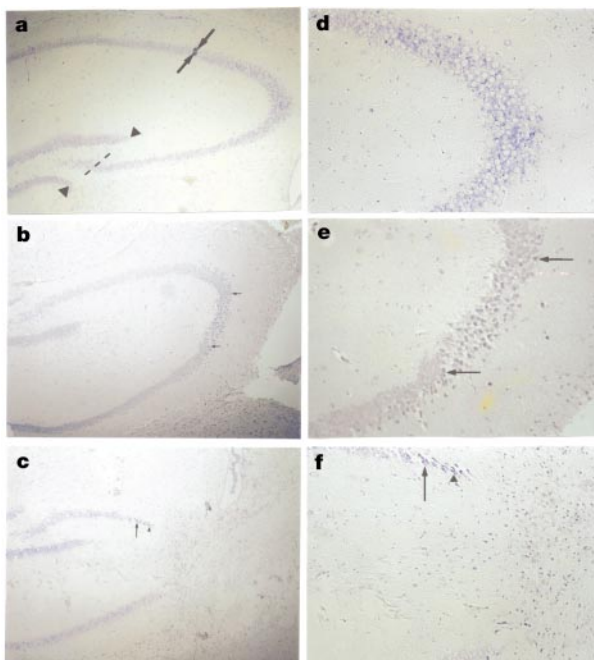


Figure 3 Hippocampal cell death after the effect of CHI is reduced by 2-AG. **a–c**, Low magnification; **d–f**, high magnification. **a, d**, The contralateral, uninjured side. The hippocampal CA3 subfield is outlined in **a**, from the CA3/2 border (arrows) up to the point at which the CA3 neurons enter the dentate gyrus (arrowheads). **b, e**, 2-AG-treated mice. Ischaemic neurons (arrows) are shown in the injured hippocampus. **c, f**, Vehicle-treated mice, after CHI. Excessive neuronal loss was the most characteristic finding, additionally to ischaemic (arrows) and necrotic neurons (arrowheads).

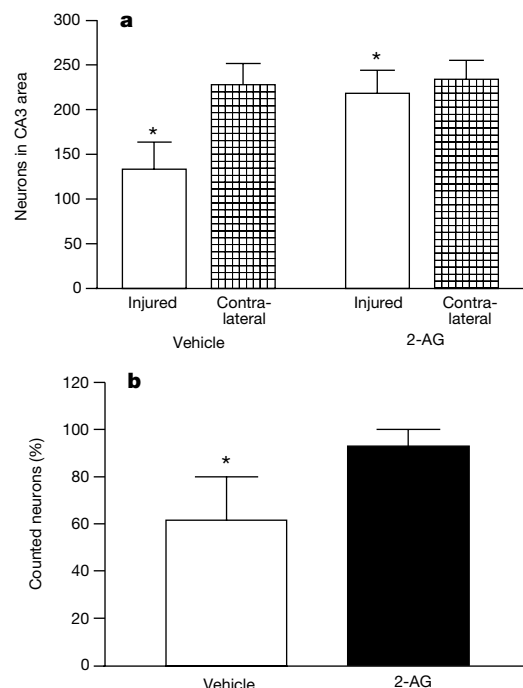


Figure 4 Effects of 2-AG treatment on the neuronal cell loss in the CA3 hippocampal subfield. **a**, The number of neurons in the injured and contralateral control sides (asterisk, $P < 0.001$). **b**, The number of neurons counted in the CA3 area in the injured hemisphere, expressed as a percentage of those counted in the CA3 area in the intact hemisphere, in vehicle-treated and 2-AG-treated mice (asterisk, $P < 0.001$). Data are mean $\pm$ s.d.

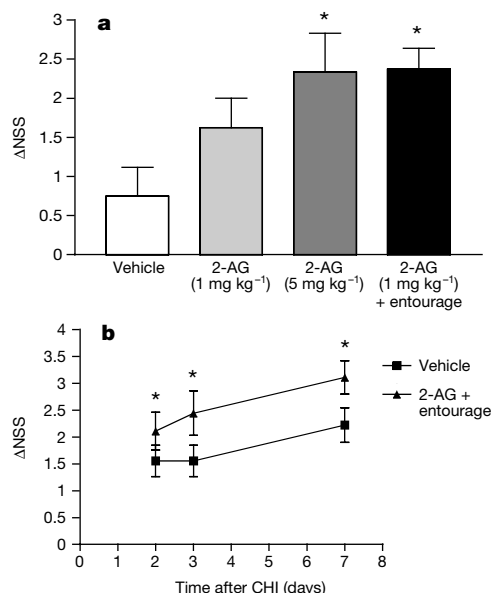


Figure 5 'Entourage' of congeners enhances the effect of 2-AG. **a**, 2-AG was injected 1 h after CHI together with its congeners, 2-palmitoyl-glycerol and 2-linoleoyl-glycerol at a dose ratio of 1, 5 and 10 mg kg⁻¹, respectively. ΔNSS was evaluated at 24 h to assess the protective effect of the mixture compared with the effect of 2-AG at the same dose. Whereas the effect of 1 mg kg⁻¹ of 2-AG by itself was not significant, in the presence of the entourage it was more effective, reaching a significant protective effect, comparable to 2-AG at a higher dose (5 mg kg⁻¹). Asterisk, $P < 0.05$ versus control CHI-traumatized mice injected with vehicle. **b**, The 2-AG plus entourage mixture was injected 1, 24 and 48 h after CHI, and its protective effects were sustained for 7 days. Data are mean ± s.e.m.

ouabain-induced *in vivo* excitotoxicity²⁵. The cytoprotective functions of *N*-acyl ethanol amines, including the endocannabinoid anandamide, have recently been reviewed²⁶. Moreover, a non-psychotropic cannabinoid analogue (HU-211), which does not bind to either CB₁ or CB₂ but which has NMDA (*N*-methyl-D-aspartate) blockade properties, has been shown to provide significant protection both *in vivo* and *in vitro*²⁷, and enhanced production of 2-AG in rat brain administered with picrotoxinin has been recorded²⁸. CB₁ agonists are also known to exert anti-inflammatory effects²⁹ and to cause hypothermia³⁰. And, as noted above⁹, 2-AG prevents formation of TNF-α and ROS. These mechanisms may contribute towards the neuroprotective profile of 2-AG.

Because cannabinoids are known to reduce body temperature⁵, and because hypothermia has been reported to provide neuroprotection, we monitored the effect of 2-AG on body temperature. No reduction of body temperature was found at 0.1 mg kg⁻¹ (a dose that abolished oedema) and only minor reduction (~1.6–1.9°C), too small to exert the observed protection, was noted after administration of 5 mg kg⁻¹. A possible mechanism by which 2-AG may protect the injured brain is related to its ability to abolish the deleterious effect of endothelin-1 (ET-1). ET-1 is the most potent vasoconstrictor known; it reduces cerebral blood flow and is important in the pathophysiology of brain trauma and ischaemia. We recently demonstrated that 2-AG counteracts ET-1-induced cerebral capillary and microvascular endothelial responses (that is, Ca²⁺ mobilization and cytoskeletal rearrangement)³¹. We suggest, therefore, that 2-AG contributes to cerebroprotection not only by reducing neuronal excitotoxicity and inhibiting production of TNF-α and ROS, but also by lowering cerebral vasoconstriction.

In response to traumatic brain injury there is local and transient accumulation of 2-AG at the site of injury, peaking at 4 h and sustained up to at least 24 h. Although the exact role of endogenous 2-AG is yet not known, the neuroprotection exerted by exogenous 2-AG suggests that it may serve as a molecular regulator of

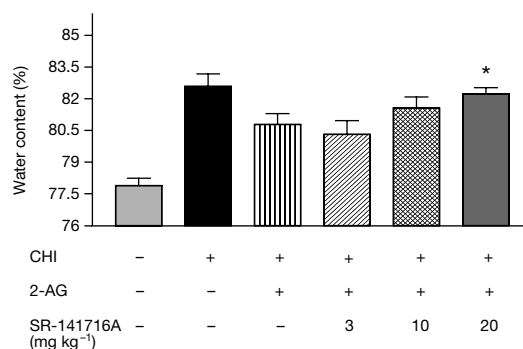


Figure 6 The protective effect of 2-AG is mediated by the CB₁ cannabinoid receptor. SR-141716A is a specific antagonist of the CB₁ brain receptor. Mice were subjected to CHI and 1 h later were injected with 2-AG (1 mg kg⁻¹) and the antagonist (at 3, 10 or 20 mg kg⁻¹). After 24 h they were killed and oedema was assessed. At its highest dose (20 mg kg⁻¹), the antagonist significantly abolished the protective effect of 2-AG, suggesting that the latter affords its protection through its action at the CB₁ receptor. Asterisk, $P < 0.05$ versus 2-AG alone. Data are mean ± s.e.m.

pathophysiological events, attenuating excitotoxicity and protecting the brain during development of secondary injury. Inhibition of this protective effect by SR-141716A suggests that 2-AG and the CB₁ receptor may be important in the pathophysiology of traumatic brain injury. However, because of the high dose of the CB₁ antagonist required to block the 2-AG protective effect, a non-CB₁-mediated mechanism of 2-AG cannot be excluded. □

Methods

Animals

The study was performed according to the guidelines of the Institutional Animal Care Committee. Male Sabra mice (Hebrew University strain) weighing 32–35 g were used. The animals were divided into groups, treated with different doses of 2-AG and killed at different times after CHI or sham surgery.

Trauma model

Trauma was induced under ether anaesthesia, which was confirmed by testing loss of pupillary and corneal reflexes, as described previously in detail¹². To assess the functional impairment after trauma, the NSS scoring system was used based on the ability of the mice to perform ten different tasks that evaluate motor ability, balancing and alertness¹³. Cerebral oedema was evaluated by determining the tissue water content in the injured brain, as previously described¹². The percentage of water in the tissue was calculated as ((wet weight – dry weight)/wet weight) × 100.

Analysis of brain 2-AG levels

Brains were frozen at the designated post-CHI times. Tissue segments were homogenized in a mixture of chloroform and methanol (2:1) containing 10 μl of 1 mM arachidin in as an internal standard. Vacuum filtration was carried out and the filter washed at least three times with the organic mixture. The solvents were evaporated to dryness. The solid residue was re-dissolved in chloroform and dried under nitrogen, and fatty acids were separated from other lipids using thin-layer chromatography (TLC) plates with fluorescent indicator. Synthetic 2-AG in chloroform was applied alongside the analysed fraction to help in the identification of endogenous 2-AG. The solvent system for developing of sample was hexane, ether, acetone and acetic acid (40:20:7:1, v/v). Using a visualizing reagent (KMnO₄), the 2-AG was identified, scraped, removed from the silica gel and placed into tubes containing chloroform and methanol (19:1); the fractions were collected and evaporated to dryness. The solid residue was dissolved in chloroform and dried under a stream of nitrogen.

GC-MS analysis

We quantitatively analysed the samples by gas chromatography–mass spectrometry (GC-MS) in a Hewlett-Packard G1800A GCD system fitted with a 5% phenylmethylsiloxane (HP-5MS) capillary column (30 m × 0.25 mm (internal diameter) × 0.25 μm), temperature programmed from 150 to 280 °C at 50 °C min⁻¹. The selected qualifier ions were (*m/z*) 74, 103 and 129 for 2-AG and 57, 73 and 147 for arachidin. Quantifier ions were 218 for 2-AG and 427 for arachidin, respectively. 2-AG amounts of 0.5–16 nmol were used for the calibration curve together with 10 nmol arachidin.

To the solid residue, which was separated by TLC, dissolved in chloroform and dried again, 10 μl of bis(trimethyl-silyl)trifluoroacetamide (BSTFA) was added to silylate the free hydroxyl groups. Two microlitres of this solution were injected into the GC-MS. 2-AG and the standard arachidin were eluted after 540 and 620 s, respectively. The ratio between 2-AG and the standard was used to calculate the level of 2-AG in the brain.

Evaluation of infarct volume

Mice were traumatized and treated with vehicle or 2-AG, and after 24 h their brains were sliced every 2 mm using a brain mould. The slices were placed in a 2% solution of 2,3,5-triphenyltetrazolium chloride (TTC) in PBS buffer¹⁴ and photographed with a Stereo-scopic Stemi SV11 (Zeiss) and digital photomicroscopy (Coolpix E990 (Nikon)). Scion Image-Release Beta 4.0.2 program was used to quantify the infarct volume.

Evaluation of hippocampal cell death

Mice were killed by perfusion fixation with 4% phosphate-buffered formaldehyde under ether anaesthesia and brains were post-fixed with the same fixative. Serial 6- μ m brain cryostat sections, cut coronally, anterior to posterior, were stained with haematoxylin and eosin and Nissl, and two uninformed observers reviewed the preparations independently. Cell counting was performed under light microscope and camera lucida on the CA3 subfield of dorsal hippocampus in both hemispheres, in 12–15 randomly chosen sections from each brain. By definition, the CA3 area started at the CA3–CA2 border up to the point where the CA3 neurons enter the dentate gyrus, as defined by the lateral aspect of the granule cell layer's dorsal and ventral leaves (Fig. 3a). Ischaemic and necrotic neurons were recognized according to morphological criteria. In an attempt to estimate the neuronal loss in the CA3 area of the injured side, the percentage of the number of neuronal cells counted in this area was evaluated against the contralateral control CA3 subfield. All neurons were included, no matter whether they were normal, ischaemic or necrotic. A paired *t*-test and *t*-tests for independent samples of groups were performed.

Received 17 May; accepted 13 August 2001.

- Povlishock, J. T. & Christman, C. W. The pathobiology of traumatically induced axonal injury in animals and humans: A review of current thoughts. *J. Neurotrauma* **12**, 555–564 (1995).
- Kochanek, P. M. *et al.* Biochemical, cellular, and molecular mechanisms in the evolution of secondary damage after severe traumatic brain injury in infants and children: Lessons learned from the bedside. *Pediatr. Crit. Care Med.* **1**, 4–19 (2000).
- Mechoulam, R. *et al.* Identification of an endogenous 2-monoglyceride, present in canine gut, that binds to cannabinoid receptors. *Biochem. Pharmacol.* **50**, 83–90 (1995).
- Sugiura, T. *et al.* 2-Arachidonoylglycerol: A possible endogenous cannabinoid receptor ligand in brain. *Biochem. Biophys. Res. Commun.* **215**, 89–97 (1995).
- Mechoulam, R., Fried, E. & Di Marzo, V. Endocannabinoids. *Eur. J. Pharmacol.* **359**, 1–18 (1998).
- Piomelli, D., Giuffrida, A., Calignano, A. & Rodriguez de Fonseca, F. The endocannabinoid system as a target for therapeutic drugs. *Trends Pharmacol. Sci.* **21**, 218–223 (2000).
- Axelrod, J. & Felder, C. C. Cannabinoid receptors and their endogenous agonist, anandamide. *Neurochem. Res.* **23**, 575–581 (1998).
- McIntosh, T. K., Juhler, M. & Wieloch, T. Novel pharmacologic strategies in the treatment of experimental traumatic brain injury. *J. Neurotrauma* **15**, 731–769 (1998).
- Gallily, R., Breuer, A. & Mechoulam, R. 2-Arachidonoylglycerol, an endogenous cannabinoid, inhibits tumor necrosis factor α production in murine macrophages, and in mice. *Eur. J. Pharmacol.* **406**, R5–R7 (2000).
- Chan, P. H. Reactive oxygen radicals in signaling and damage in the ischemic brain. *J. Cereb. Blood Flow Metab.* **21**, 2–14 (2001).
- Shohami, E., Ginis, I. & Hallenbeck, J. M. Dual role of tumor necrosis factor α in brain injury. *Cytokines Growth Factor Rev.* **10**, 119–130 (1999).
- Chen, Y., Constantini, S., Trembovler, V., Weinstock, M. & Shohami, E. An experimental model of closed head injury in mice: Pathophysiology, histopathology and cognitive deficits. *J. Neurotrauma* **13**, 557–568 (1996).
- Beni-Adani, L. *et al.* A peptide derived from activity-dependent neuroprotective protein (ADNP) ameliorates injury response in closed head injury in mice. *J. Pharmacol. Exp. Ther.* **296**, 57–63 (2001).
- Mathews, K. S. *et al.* Rapid quantification of ischaemic injury and cerebroprotection in brain slices using densitometric assessment of 2,3,5-triphenyltetrazolium chloride staining. *J. Neurosci. Meth.* **102**, 43–51 (2000).
- Ben-Shabat, S. *et al.* An entourage effect: Inactive endogenous fatty acid glycerol esters enhance 2-arachidonoyl glycerol cannabinoid activity. *Eur. J. Pharmacol.* **353**, 23–31 (1998).
- Adams, I. E., Compton, D. R. & Martin, B. R. Assessment of anandamide interaction with the cannabinoid brain receptor: SR141716A antagonism studies in mice and autoradiographic analysis of receptor binding in rat brain. *J. Pharmacol. Exp. Ther.* **284**, 1209–1217 (1998).
- Smith, S. R., Terminelli, C. & Denhardt, G. Effects of cannabinoid receptor agonist and antagonist ligands on production of inflammatory cytokines and anti-inflammatory interleukin-10 in endotoxemic mice. *J. Pharmacol. Exp. Ther.* **293**, 136–150 (2000).
- Shohami, E., Shapira, Y., Yadid, G., Reisfeld, N. & Yedgar, S. Brain phospholipase A2 is activated after experimental closed head injury in rats. *J. Neurochem.* **53**, 1541–1546 (1989).
- Nadler, V., Biegon, A., Beit-Yannai, E., Adamchik, J. & Shohami, E. ⁴⁵Ca accumulation in rat brain after closed head injury: attenuation by the novel neuroprotective agent HU-211. *Brain Res.* **685**, 1–11 (1995).
- Reggio, P. H. & Traore, H. Conformational requirements for endocannabinoid interaction with the cannabinoid receptors, the anandamide transporter and fatty acid amidohydrolase. *Chem. Phys. Lipids* **108**, 15–35 (2000).
- Goparaju, S. K., Ueda, N., Yamaguchi, H. & Yamamoto, S. Anandamide amidohydrolase reacting with 2-arachidonoylglycerol, another cannabinoid receptor ligand. *FEBS Lett.* **422**, 69–73 (1998).
- Sinor, A. D., Irvin, S. M. & Greenberg, D. A. Endocannabinoids protect cerebral cortical neurons from *in vitro* ischemia in rats. *Neurosci. Lett.* **278**, 157–160 (2000).
- Shen, M. & Thayer, S. A. Cannabinoid receptor agonists protect cultured rat hippocampal neurons from excitotoxicity. *Mol. Pharmacol.* **54**, 459–462 (1998).
- Nagayama, T. *et al.* Cannabinoids and neuroprotection in global and focal cerebral ischemia and in neuronal cultures. *J. Neurosci.* **19**, 2987–2995 (1999).
- van der Stelt, M. *et al.* Neuroprotection by Delta9-tetrahydrocannabinol, the main active compound in marijuana, against ouabain-induced *in vivo* excitotoxicity. *J. Neurosci.* **21**, 6475–6479 (2001).
- Hansen, H. S., Moesgaard, B., Hansen, H. H. & Petersen, G. N-acyl ethanolamines and precursor phospholipids—relation to cell injury. *Chem. Phys. Lipids* **108**, 135–150 (2000).

- Shohami, E. & Mechoulam, R. Dexanabinol (HU-211): A nonpsychotropic cannabinoid with neuroprotective properties. *Drug Develop. Res.* **50**, 211–215 (2000).
- Sugiura, T., Yoshinaga, N., Kondo, S., Waku, K. & Ishima, Y. Generation of 2-arachidonoylglycerol, an endogenous cannabinoid receptor ligand, in picrotoxinin-administered rat brain. *Biochem. Biophys. Res. Commun.* **271**, 654–658 (2000).
- Klein, T. W., Lane, B., Newton, C. A. & Friedman, H. The cannabinoid system and cytokine network. *Proc. Soc. Exp. Biol. Med.* **225**, 1–8 (2000).
- Ovadia, H., Wohlman, A., Mechoulam, R. & Weidenfeld, J. Characterization of the hypothermic effect of the synthetic cannabinoid HU-210 in the rat. Relation to the adrenergic system and endogenous pyrogens. *Neuropharmacology* **34**, 175–180 (1995).
- Chen, Y. *et al.* Human brain capillary endothelium. 2-Arachidonoylglycerol (endocannabinoid) interacts with endothelin-1. *Circ. Res.* **87**, 323–327 (2000).

Acknowledgements

We thank the US National Institute of Drug Abuse and the Israel Science Foundation for support (grants to R.M.). E.S. and R.M. are affiliated with the David R. Bloom Center for Pharmacy at the Hebrew University of Jerusalem's School of Pharmacy.

Correspondence and requests for materials should be addressed to E.S. (e-mail: esty@cc.huji.ac.il).

A synthetic glycolipid prevents autoimmune encephalomyelitis by inducing T_H2 bias of natural killer T cells

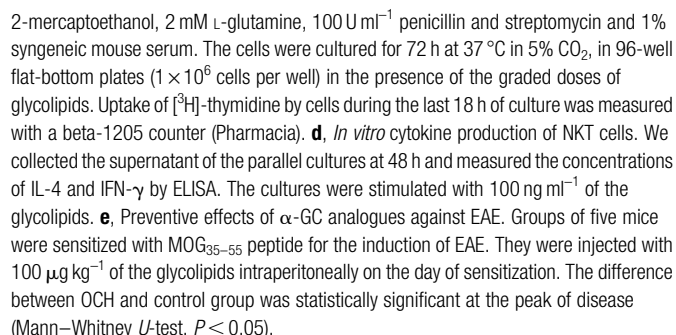
Katsuichi Miyamoto*, Sachiko Miyake* & Takashi Yamamura

Department of Immunology, National Institute of Neuroscience, NCNP, 4-1-1 Ogawahigashi, Kodaira, Tokyo 187-8502, Japan

* These authors contributed equally to this work

Experimental autoimmune encephalomyelitis (EAE) is a prototype autoimmune disease mediated by type 1 helper T (T_H1) cells and under the control of regulatory cells^{1–3}. Here we report that a synthetic glycolipid ligand for CD1d-restricted natural killer T (NKT) cells expressing the semi-invariant T-cell receptor (V α 14⁺) is preventive against EAE. The ligand is an analogue of α -galactosylceramide (α -GC), a prototype NKT cell ligand, with a truncated sphingosine chain. α -GC causes NKT cells to produce both interferon (IFN)- γ and interleukin (IL)-4 (refs 4, 5). However, this new ligand can induce a predominant production of IL-4 by the NKT cells. A single injection of this glycolipid, but not of α -GC, consistently induced T_H2 bias of autoimmune T cells by causing NKT cells to produce IL-4, leading to suppression of EAE. The lack of polymorphism of CD1d and cross-reactive response of mouse and human NKT cells to the same ligand⁶ indicates that targeting NKT cells with this ligand may be an attractive means for intervening in human autoimmune diseases such as multiple sclerosis.

Conventional T cells respond to peptide ligands in the context of the major histocompatibility complex (MHC). In contrast, the CD1d-restricted V α 14⁺ NKT cells, which express NK cell markers and the T-cell receptor (TCR) with the invariant α -chain^{7–9}, react to the glycolipid ligand bound to the monomorphic CD1d molecule and produce large amounts of IFN- γ and IL-4 within hours of TCR ligation. Although the natural ligand for NKT cells is unknown, a derivative of marine sponge, α -GC^{4,5}, is now established as a prototype NKT cell ligand inducing proliferation, cytokine production^{4,5} and apoptotic death¹⁰ of NKT cells. According to the trimolecular model of CD1–glycolipid antigen interactions, the TCR of NKT cells is predicted to contact the hydrophilic galactose moiety of α -GC, whereas two aliphatic hydrocarbon chains would



EAE induced in wild-type C57BL/6 (B6) mice¹⁷. This is because NKT-cell-derived IFN- γ would mask the therapeutic effect of the IL-4 simultaneously produced by NKT cells¹⁷. A peptide in which one or two TCR contact residues have been changed (an altered peptide ligand, APL) may induce only a subset of the functional responses seen with the unaltered peptide¹⁸. Given this information, we speculated that an altered form of α -GC might selectively

B6 mice or NKT-cell-deficient mice^a were sensitized with MOG_{35–55} for induction of EAE. Control PBS solution or 400 µg kg⁻¹ of α-GC or OCH diluted in PBS was orally injected with a cannula on the day of immunization (day 0) or on day 8, when the histopathological signs of EAE could be seen in most mice. Data shown are mean ± s.e.m.

^a *P* < 0.05 (Mann–Whitney *U*-test, OCH versus control).

stimulate the production of IL-4 and exhibit protective effects against EAE. To verify this idea, we synthesized a panel of α -GC analogues by replacing the sugar moiety and/or truncating the aliphatic chains (Fig. 1a).

To test our prediction of a therapeutic effect on EAE, we first measured the serum levels of IFN- γ and IL-4 after intraperitoneal injection of the glycolipids into B6 mice (Fig. 1b). In agreement with previous results¹⁷, injection of α -GC induced a rapid elevation of IL-4 with the peak value at 2 h and a delayed and prolonged elevation of IFN- γ in B6 mice. However, NKT-cell-deficient TCR $\text{J}\alpha 281^{-/-}$ mice⁴ did not respond to the glycolipid (data not shown), indicating that elevation of the cytokines in B6 mice was mediated by NKT cells. Although the NH analogue was non-stimulatory as judged by this *in vivo* assay, the 3,4D and OCH analogues were regarded as partial agonists compared with α -GC. Stimulation with 3,4D induced a lower production of IL-4 and IFN- γ than did α -GC injection, whereas OCH injection dissociated the production of IL-4 and IFN- γ : production of IL-4 was unaffected but IFN- γ was much lower.

We then compared the ability of α -GC and its analogues to induce cell proliferation and cytokine production by NKT cells *in vitro*. NH and 3,4D analogues were non-stimulatory in these assays, whereas α -GC and OCH induced both cell proliferation and cytokine responses. Notably, compared with α -GC, OCH was about five- to tenfold less active in inducing cell proliferation (Fig. 1c) and induced less IFN- γ but more IL-4 *in vitro* (Fig. 1d). Together with the *in vivo* results (Fig. 1b), we postulate that OCH may induce a T_H2 bias through a predominant production of IL-4 by NKT cells and thus could be useful as a therapeutic agent for T_H1 -mediated diseases.

To find support for this postulate, we examined the effects of α -GC and the analogues on EAE induced by active sensitization against a peptide comprising amino acids 35–55 of the myelin oligodendrocyte glycoprotein (MOG_{35–55}). Whereas α -GC and NH did not inhibit development of EAE (Fig. 1e), a single intraperitoneal injection of OCH significantly inhibited the clinical signs of EAE. 3,4D analogue also tended to suppress EAE, but the suppression was not significant. We next administered α -GC or OCH orally and induced EAE. The mice given oral OCH on the day of immunization (day 0) again showed lower clinical and pathological grades than the control mice given PBS or α -GC (Table 1a). However, induction of EAE in NKT-cell-deficient TCR $\text{J}\alpha 281^{-/-}$ mice was not altered by oral administration of OCH or α -GC (Table 1c), indicating that NKT cells are involved in the OCH-induced suppression of EAE. We next administered OCH or α -GC orally on day 8 after immunization. At this point, histopathological signs of EAE could be verified in most mice and a small proportion of mice started to develop the first clinical sign, flaccid tail. Notably, this 'day 8 injection' protocol was found to be effective in reducing the cumulative disease score as well as the maximum clinical score (Table 1b). T_H1 -mediated EAE was therefore significantly sup-

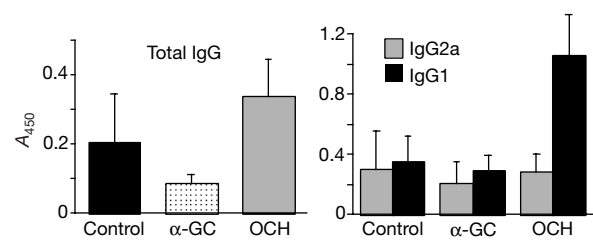


Figure 2 Elevation of IgG1 isotype of anti-MOG_{35–55} antibody in the OCH-treated mice. A₄₅₀ is the absorbance at wavelength 450 nm. Groups of six mice were sensitized with MOG_{35–55} peptide and fed orally 400 $\mu\text{g kg}^{-1}$ of α -GC or OCH with a cannula; the serum samples were collected on day 30. Different dilutions of the sera were assayed for total IgG, IgG1 and IgG2a of anti-MOG_{35–55} antibody by ELISA¹⁷. Shown are the representative values obtained from the sera diluted at 1 : 10.

pressed by OCH, which induces a predominant production of IL-4 by NKT cells.

To confirm the involvement of IL-4 in the suppression of EAE, we next examined whether the preventive effect of OCH is abrogated after neutralization of IL-4 *in vivo*. Groups of mice were injected with or without anti-IL-4 antibody and given α -GC or OCH orally. Remarkably, the preventive effect of OCH against EAE was no more evident when anti-IL-4 was injected, although injection of this antibody did not modulate EAE in the control mice (Table 2). These results imply that IL-4 is critical in the OCH-mediated suppression of EAE and are consistent with our postulate that this compound modulates EAE by stimulating IL-4 production by NKT cells.

We also measured serum levels of immunoglobulin- $\gamma 1$ (IgG1) and IgG2a isotypes of anti-MOG_{35–55} antibodies because an isotype shift toward IgG1 indicates a deviation of T cells toward T_H2 . The total IgG and IgG1 titres against MOG_{35–55} were elevated but IgG2a was unchanged in the mice treated with OCH (Fig. 2), indicating a T_H2 bias of MOG_{35–55}-reactive T cells *in vivo*. This observation further supports that IL-4 mediates the suppression of EAE by OCH.

As seen with APL¹⁸, the altered glycolipid ligands of α -GC triggered different responses of cytokine production and cell proliferation. However, the experiments shown above did not clarify whether the effect of OCH represents a true qualitative difference between this compound and α -GC. In fact, because OCH is less active than α -GC in inducing the proliferation of NKT cells (Fig. 1c), it is possible that α -GC, given at lower doses, might induce selective production of IL-4 by NKT cells and exhibit an OCH-like immunomodulatory effect. To address this issue, we injected different doses of α -GC into B6 mice and examined the serum levels of IFN- γ and IL-4 (Fig. 3). Notably, the relative ratio of serum IFN- γ to IL-4 at their peak was not significantly altered by reducing the dose of α -GC. In addition, injection of lower doses of α -GC did not inhibit the development of EAE (data not shown). These results imply that OCH is not simply a weaker agonist than α -GC but a qualitatively different ligand for NKT cells.

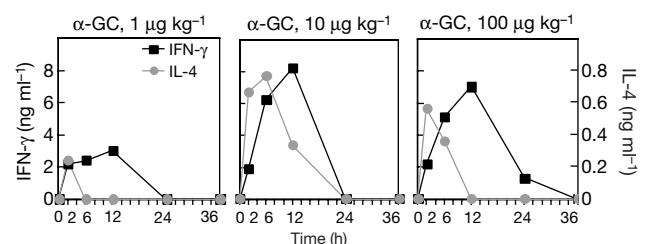


Figure 3 Serum cytokine kinetics after injection of different doses of α -GC. Pairs of B6 mice were injected intraperitoneally with 1, 10 or 100 $\mu\text{g kg}^{-1}$ of α -GC, and the levels of IL-4 and IFN- γ in the pooled sera were determined at the indicated points. This is a representative experiment of two with similar results.

Table 2 Suppression of EAE by oral OCH analogue is abrogated after neutralizing IL-4 *in vivo*

	Maximum score	Day of EAE onset	Incidence	Cumulative clinical score
Without anti-IL-4				
Control	3.78 $\pm$ 0.24	9.67 $\pm$ 0.97	10/10	39.94 $\pm$ 3.65
α -GC	4.00 $\pm$ 0.25	9.10 $\pm$ 0.84	10/10	45.00 $\pm$ 3.87
OCH	2.75 $\pm$ 0.37*	11.00 $\pm$ 0.78	9/10	26.05 $\pm$ 4.16*
With anti-IL-4				
Control	4.00 $\pm$ 0.29	8.57 $\pm$ 0.84	10/10	42.50 $\pm$ 3.40
α -GC	3.79 $\pm$ 0.29	11.29 $\pm$ 1.58	10/10	37.00 $\pm$ 5.02
OCH	3.50 $\pm$ 0.15	9.43 $\pm$ 0.90	10/10	38.79 $\pm$ 2.94

B6 mice were sensitized with MOG_{35–55} for induction of EAE. Control PBS solution or 400 $\mu\text{g kg}^{-1}$ of α -GC or OCH diluted in PBS was orally injected with a cannula. Three groups of mice were injected intraperitoneally with a neutralizing anti-IL-4 monoclonal antibody (11B11, 1 mg per mouse); the other three groups were injected with PBS. Data shown are mean $\pm$ s.e.m.

* $P < 0.05$ (Mann-Whitney U-test, OCH versus control).

The OCH analogue has shorter hydrophobic chains, particularly for the sphingosine base, but shares the same hydrophilic cap with α -GC. Although the mechanism by which OCH induces T_H2 bias of NKT cells remains to be investigated, we hypothesize that a less tight binding of OCH than of α -GC to CD1d may make the orientation of the galactose hydrophilic cap unstable. This is likely to lead to the qualitatively different interactions of OCH with the TCR, resulting in an altered T_H1/T_H2 balance.

Although both rodent and human NKT cells recognize α -GC in the context of CD1d^{5,6}, we have recently found that human NKT cells also recognize OCH (M. Araki *et al.*, unpublished data). In addition, whereas α -GC induces the apoptotic cell death of NKT cells¹⁰, OCH increases the number of NKT cells in the liver (data not shown). Weekly injections of OCH (four times) during the recovery phase of EAE did not induce anaphylactic shock in any of the mice, whereas such allergic reactions are deleterious consequences of a T_H2 shift in conventional T cells by treatment with peptide^{19,20}. These properties encourage us to apply OCH or related ligands for prevention and treatment of human autoimmune diseases that are associated with a reduction in and/or a T_H1 bias of NKT cells^{14–16}. □

Methods

Glycolipids

The glycolipids were dissolved in PBS buffer containing 10% dimethylsulphoxide (DMSO) at 100 $\mu\text{g ml}^{-1}$ and kept at -20°C until used. For study, the stock solution was further diluted with PBS. PBS solution was supplemented with an equivalent amount of DMSO to provide a control for the experiments.

Induction and assessment of EAE

B6 mice, maintained under specific pathogen-free conditions, were injected in the tail base bilaterally with 200 μl of an inoculum containing 100 μg of MOG_{35–55} (amino acids MEVGWYRSPFSRVVHLYRNGK) and 1 mg of *Mycobacterium tuberculosis* in incomplete Freund adjuvant. Pertussis toxin (200 ng) was injected intravenously on day 0 and day 2 after immunization. The clinical signs were scored as follows: 0, no clinical signs; 1, partial loss of tail tonicity; 2, completely limp tail and abnormal gait; 3, partial hindlimb paraparesis; 4, complete hindlimb paraparesis; 5, fore- and hindlimb paralysis or a moribund state. On day 21 after immunization, several mice were randomly chosen from each group and examined for histopathological grades of EAE. Briefly, mice were perfused with 0.01 M PBS first and then with a fixative containing 4% paraformaldehyde and 0.2% picric acid in 0.1 M PBS. Paraffin sections of brain and spinal cords (6 μm thickness) were stained with haematoxylin and eosin and Klüver–Barrera methods. The histopathological grades of EAE were scored as follows: 1, leptomeningeal and adjacent subpial cell infiltration; 2, mild perivascular cuffing; 3, extensive perivascular cuffing; 4, extensive periventricular cuffing with severe parenchymal cell infiltration.

Received 5 July; accepted 3 August 2001.

- Nicholson, L. B. & Kuchroo, V. K. Manipulation of the T_H1/T_H2 balance in autoimmune disease. *Curr. Opin. Immunol.* **8**, 837–842 (1996).
- Kumar, V., Aziz, F., Sercarz, E. & Miller, A. Regulatory T cells specific for the same framework 3 region of the V β 8.2 chain are involved in the control of collagen II-induced arthritis and experimental autoimmune encephalomyelitis. *J. Exp. Med.* **185**, 1725–1733 (1997).
- Zhang, B.-n., Yamamura, T., Kondo, T., Fujiwara, M. & Tabira, T. Regulation of experimental autoimmune encephalomyelitis by natural killer (NK) cells. *J. Exp. Med.* **186**, 1677–1687 (1997).
- Kawano, T. *et al.* CD1d-restricted and TCR-mediated activation of V α 14 NKT cells by glycosylceramides. *Science* **278**, 1626–1629 (1997).
- Spada, F. M., Koezuka, Y. & Porcelli, S. A. CD1d-restricted recognition of synthetic glycolipid antigens by human natural killer T cells. *J. Exp. Med.* **188**, 1529–1534 (1998).
- Brossay, L. *et al.* CD1d-mediated recognition of α -galactosylceramide by natural killer T cells is highly conserved through mammalian evolution. *J. Exp. Med.* **188**, 1521–1528 (1998).
- Bix, M. & Locksley, R. M. Natural T cells. Cells that co-express NKR1-1 and TCR. *J. Immunol.* **155**, 1020–1022 (1995).
- Bendelac, A., Rivera, M. N., Park, S. & Roark, J. H. Mouse CD1-specific NK1 T cells: development, specificity and function. *Annu. Rev. Immunol.* **15**, 535–562 (1997).
- Godfrey, D. I., Hammond, K. J., Poulton, L. D., Smyth, M. J. & Baxter, A. G. NKT cells: Facts, functions and fallacies. *Immunol. Today* **21**, 573–583 (2000).
- Matsuda, J. L. *et al.* Tracking the response of natural killer T cells to a glycolipid antigen using CD1d tetramer. *J. Exp. Med.* **192**, 741–754 (2000).
- Moody, D. B., Besra, G. S., Wilson, I. A. & Porcelli, S. A. The molecular basis of CD1-mediated presentation of lipid antigens. *Immunol. Rev.* **172**, 285–296 (1999).
- Mieza, M. A. *et al.* Selective reduction of V α 14⁺ NK T cells associated with disease development in autoimmune-prone mice. *J. Immunol.* **156**, 4035–4040 (1996).
- Kirsten, J. L. *et al.* α/β T cell receptor (TCR)⁺ CD4⁺ CD8⁺ (NKT) thymocytes prevent insulin-dependent diabetes mellitus in nonobese diabetic (NOD)/Lt mice by the influence of interleukin (IL)-4 and/or IL-10. *J. Exp. Med.* **187**, 1047–1056 (1998).
- Sumida, T. *et al.* Selective reduction of T cells bearing invariant V α 24J α Q antigen receptor in patients with systemic sclerosis. *J. Exp. Med.* **182**, 1163–1168 (1995).

- Wilson, S. B. *et al.* Extreme T_H1 bias of invariant V α 24J α Q T cells in type 1 diabetes. *Nature* **391**, 177–181 (1998).
- Illés, Zs. *et al.* Differential expression of natural killer T cell V α 24J α Q invariant TCR chain in the lesions of multiple sclerosis and chronic inflammatory demyelinating polyneuropathy. *J. Immunol.* **164**, 4375–4381 (2000).
- Pál, E. *et al.* Costimulation-dependent modulation of experimental autoimmune encephalomyelitis by ligand stimulation of V α 14 NK T cells. *J. Immunol.* **166**, 662–668 (2001).
- Sloan-Lancaster, J. & Allen, P. M. Altered peptide ligand-induced partial T cell activation: molecular mechanisms and role in T cell biology. *Annu. Rev. Immunol.* **14**, 1–27 (1996).
- Kappos, L. *et al.* Induction of a non-encephalitogenic type 2 T helper-cell autoimmune response in multiple sclerosis after administration of an altered peptide ligand in a placebo-controlled, randomized phase II trial. *Nature Med.* **6**, 1176–1182 (2000).
- Pedotti, R. *et al.* An unexpected version of horror autotoxicus: Anaphylactic shock to a self-peptide. *Nature Immunol.* **2**, 216–222 (2001).
- Morita, M. *et al.* Synthesis of α -, β -monoglycosylceramides and four diastereomers of an α -galactosylceramide. *Bioorg. Med. Chem. Lett.* **5**, 699–704 (1995).

Acknowledgements

We thank M. Taniguchi for supplying TCR J α 281^{–/–} mice. This work was supported by a Research on Brain Science grant from the Ministry of Health, Labor and Welfare in Japan.

Correspondence and requests for materials should be addressed to T.Y. (e-mail: yamamura@ncnp.go.jp).

Direct ligand–receptor complex interaction controls *Brassica* self-incompatibility

Seiji Takayama^{*†}, Hiroko Shimosato^{*†}, Hiroshi Shiba[†], Miyuki Funato[†], Fang-Sik Che[†], Masao Watanabe[‡], Megumi Iwano[†] & Akira Isogai[†]

[†] Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma 630-0101, Japan

[‡] Faculty of Agriculture, Iwate University, 3-18-8 Ueda, Morioka 020-8550, Japan

* These authors contributed equally to this work

Many higher plants have evolved self-incompatibility mechanisms to prevent self-fertilization¹. In *Brassica* self-incompatibility, recognition between pollen and the stigma is controlled by the *S* locus, which contains three highly polymorphic genes: *S*-receptor kinase (*SRK*)², *S*-locus protein 11 (*SP11*)³ (also called *S*-locus cysteine-rich protein; *SCR*)⁴ and *S*-locus glycoprotein (*SLG*)⁵. *SRK* encodes a membrane-spanning serine/threonine kinase that determines the *S*-haplotype specificity of the stigma⁶, and *SP11* encodes a small cysteine-rich protein that determines the *S*-haplotype specificity of pollen^{4,7,8}. *SP11* is localized in the pollen coat⁸. It is thought that, during self-pollination, *SP11* is secreted from the pollen coat and interacts with its cognate *SRK* in the papilla cell of the stigma to elicit the self-incompatibility response. *SLG* is a secreted stigma protein⁹ that is highly homologous to the *SRK* extracellular domain. Although it is not required for *S*-haplotype specificity of the stigma, *SLG* enhances the self-incompatibility response⁶; however, how this is accomplished remains controversial^{10–12}. Here we show that a single form of *SP11* of the *S*₈ haplotype (*S*₈-*SP11*) stabilized with four intramolecular disulphide bonds specifically binds the stigma membrane of the *S*₈ haplotype to induce autophosphorylation of *SRK*₈, and that *SRK*₈ and *SLG*₈ together form a high-affinity receptor complex for *S*₈-*SP11* on the stigma membrane.

To examine whether *SP11* interacts with *SRK* and/or *SLG* of the same *S* haplotype, we first purified *S*₈-*SP11* from pollen coat using an affinity-purified *S*₈-*SP11* antibody⁸. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) analysis of the immunoprecipitates determined the presence of *S*₈-*SP11* with a relative molecular mass (*M*_r) of 5,716 in the

pollen extract of the S_8 haplotype but not in that of the S_9 haplotype (Fig. 1a). This M_r value is consistent with that calculated on the basis of the following assumptions: first, S_8 -SP11 is present as a monomer; second, the predicted signal peptide is removed at the expected cleavage site^{4,7}; and last, all eight cysteine residues are oxidized to form four intramolecular disulphide bonds.

To characterize S_8 -SP11 further, we chemically synthesized S_8 -SP11 by both solid-phase and solution methods. When synthetic S_8 -SP11 was reduced and then subjected to mild oxidation, only one main oxidized form was obtained, with all eight cysteine residues involved in intramolecular disulphide bonds. This oxidized form of S_8 -SP11 exhibits the same MS spectrum as does S_8 -SP11 purified from the pollen coat (Fig. 1a). The disulphide linkage (Fig. 1b) was determined by analysing proteolytic fragments of the synthetic S_8 -SP11 using Edman degradation and MS spectrometry. The disulphide bond arrangement of S_8 -SP11 (C1–C8, C2–C5, C3–C6, C4–C7) is identical to that of plant defensin family proteins¹³—antimicrobial small cysteine-rich proteins represented by γ 1-P (ref. 14)—even though the positioning of the cysteine residues, especially the C4 residue, is different^{3,4,7,15}.

We next examined whether the principal oxidized form of SP11 might function as the pollen S -haplotype determinant of self-incompatibility. We previously used a pollen hydration bioassay system¹⁶ to show that applying S_9 -SP11 (containing a mixture of varying degrees of oxidized forms) expressed from *Escherichia coli* to the papilla cells of S_9 stigmas, but not S_8 stigmas, results in the inhibition of hydration of cross-pollen that is subsequently placed on the stigma surface⁷.

Because this bioassay requires the use of a micromanipulator and much skill, we have developed a simpler method. In our revised system, the oxidized form of SP11 is directly applied to the whole surface of the stigma using a micropipette and the stigma is pollinated with cross-pollen. Compatibility of the pollination is

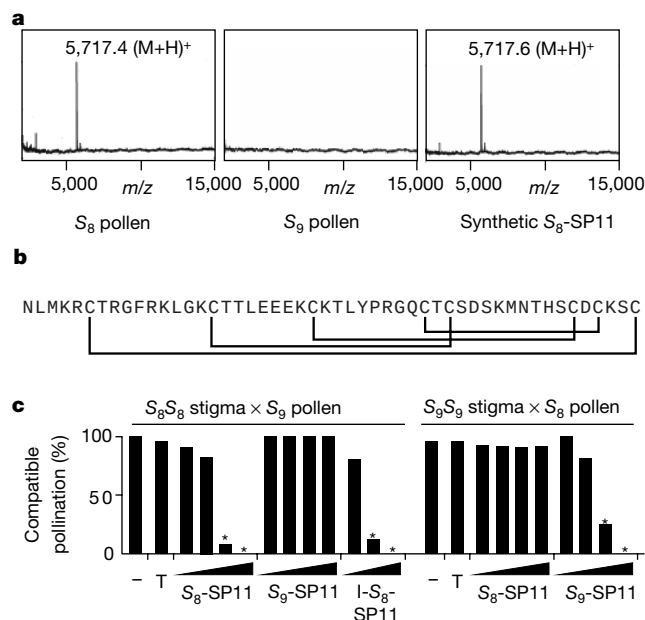


Figure 1 Molecular characterization of S_8 -SP11. **a**, MALDI-TOF-MS spectra of proteins immunoprecipitated from pollen extracts of S_8 and S_9 haplotypes, and synthetic S_8 -SP11. **b**, Primary structure of S_8 -SP11. The four intramolecular disulphide linkages are indicated by connecting lines. **c**, Pollination bioassay for SP11 proteins. Effect of pretreatments of stigma with S_8 -SP11 or S_9 -SP11 (each at 0.5, 5, 50 or 500 fmol per stigma) or non-radioactive iodinated S_8 -SP11 (I- S_8 -SP11, at 5, 50 or 500 fmol per stigma). —, no treatment; T, 0.05% Tween-20-treated control. Data represent the percentage of compatible pollination ($n = 11$ –28 experiments; asterisk, $P < 0.05$ with Fisher's exact probability test versus the Tween-20-treated control).

assessed by the penetration of pollen tubes into each stigma. In addition to the oxidized form of the synthetic S_8 -SP11, a recombinant S_9 -SP11 (ref. 7) was also oxidized, purified to a single form and used for the bioassay. Both SP11 proteins inhibited the penetration of compatible pollen tubes in an S -haplotype-specific and dose-dependent manner (Fig. 1c). Because the synthetic S_8 -SP11 appears to be biologically active, and because of the difficulty in obtaining sufficient amounts of purified S_8 -SP11, we used the synthetic form of S_8 -SP11 in all further experiments.

In the bioassay system described above, the oxidized form of an SP11 molecule might potentially interact with its cognate SRK and/or SLG to induce an incompatible response in papilla cells, and thereby inhibit penetration of cross-pollen tubes. To address this possibility, we first used the BIAcore system to examine whether SP11 would interact with SLG. This system has been effective in monitoring the interactions between a defensin-like pollen coat protein (PCP) family of proteins, represented by PCP-A1 (ref. 17), and an SLG family of proteins¹⁸. Both S_8 -SP11 and S_9 -SP11, even at 100 nM, showed little or no affinity for either SLG₈ or SLG₉ (Fig. 2a).

To investigate whether SP11 interacts with SRK, we produced a recombinant extracellular domain (S domain) of SRK₉ linked to glutathione S -transferase (SRK₉-SD–GST) in silkworm larvae using a recombinant baculovirus. S-locus-related protein 1 (SLR1), which is a homologue of SLG and interacts with a PCP family protein, SLR1-BP1 (SLR1-binding protein 1)¹⁸, was similarly expressed as a GST fusion protein (SLR1–GST) and used as a control. SLR1–GST showed a high affinity for SLR1-BP1, as did purified SLR1, suggesting that this expression system can express an SLG family protein in a physiologically active form (Fig. 2a, b). However, SRK₉-SD–GST, like SLG₈ and SLG₉, showed no affinity for either S_8 -SP11 or S_9 -SP11 (Fig. 2b). At present, we cannot completely rule out the possibility that the lack of interaction in the latter case was due to incorrect folding or post-translational modification of SRK₉-SD–GST produced in a heterologous expression system; however, our results are consistent with findings that an active kinase domain of two leucine-rich-repeat receptor kinases of *Arabidopsis*, CLAVATA1 and FLS2, is essential for the assembly of the receptor complex and for the ligand binding^{19,20}.

To identify the stigmatic receptor(s) for SP11, we first chemically

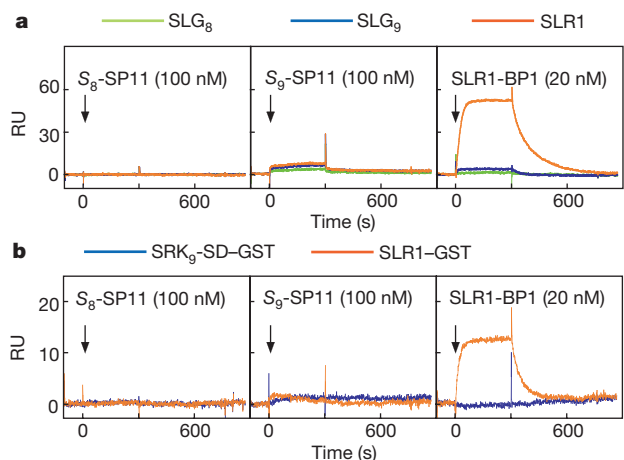


Figure 2 Surface plasmon resonance measurement of the affinity of SP11s for SLGs and a recombinant SRK extracellular domain. **a**, Purified stigmatic SLG₈, SLG₉ and SLR1 were separately coupled to parallel flow cells of a CM5 sensor chip¹⁸, and the indicated concentrations of S_8 -SP11, S_9 -SP11 and SLR1-BP1 were injected for 0–300 s. **b**, Recombinant GST fusion proteins of SRK extracellular domain (SRK₉-SD–GST) and of SLR1 (SLR1–GST) were separately coupled to a sensor chip through an anti-GST antibody. A flow cell coupled with GST was used as a control for background subtracting. RU, resonance units.

modified S_8 -SP11 with ^{125}I or non-radioactive iodine. The high performance liquid chromatography (HPLC)-purified non-radioactive iodinated S_8 -SP11 (I- S_8 -SP11) has similar biological activity to S_8 -SP11 (Fig. 1c). ^{125}I -labelled S_8 -SP11 specifically binds the stigmatic microsomal membranes of the S_8 homozygote (Fig. 3a). Scatchard analysis indicates that there is a high-affinity binding site (dissociation constant (K_d) = 1.2 nM; maximum number of binding (B_{max}) = 250 fmol per mg membrane protein) and a low-affinity binding site (K_d = 32 nM; B_{max} = 1 pmol per mg membrane protein) (Fig. 3b). As both high- and low-binding sites are present specifically in the membrane of the cognate S haplotype, they both are expected to comprise S -locus protein(s); however, the molecular basis for the binding difference remains to be determined. The binding affinity at the high-affinity site is stronger than that of the BRI1-brassinolide interaction (K_d = 7.4–10.8 nM)—the only other plant receptor-kinase–ligand interaction that has been analysed stoichiometrically²¹.

To identify the receptor components responsible for the interaction of stigmatic microsomal membranes of the S_8 homozygote with ^{125}I -labelled S_8 -SP11, we performed crosslinking experiments using the chemical bis[sulphosuccinimidyl] suberate (BS^3). Two protein bands of M_r 120,000 (120K) and 65,000 (65K) were detected by SDS polyacrylamide gel electrophoresis (SDS–PAGE); neither protein band was detected when a 100-fold molar excess of S_8 -SP11 was present during the crosslinking (Fig. 4a). Moreover, these two protein bands were not detected when microsomal membranes of the S_9 homozygote were used in crosslinking (Fig. 4a). Together, these results indicate that there may be two receptor components for S_8 -SP11 in the stigma of the S_8 homozygote.

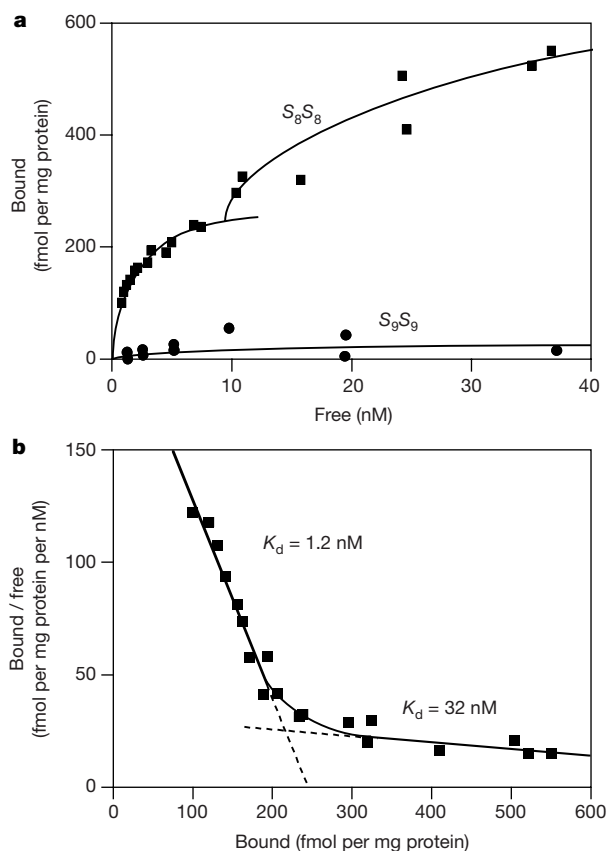


Figure 3 Specific ^{125}I -labelled S_8 -SP11 binding to microsomal membranes of the stigma. **a**, Specific binding was determined by subtracting the binding in the presence of 100-fold S_8 -SP11 from the total binding in the absence of the cold competitor. Representative data from one of three repeat experiments are shown. **b**, Scatchard plot of the binding data shown in **a**.

As the intensity of both protein bands increased in parallel when increased concentrations of ^{125}I -labelled S_8 -SP11 (0.1–10 nM) were used in crosslinking, these two proteins probably form one single high-affinity binding site (Fig. 4b). The molecular masses of these two protein bands suggest that one could be SRK₈ and the other SLG₈. To confirm the identity, we solubilized chemically crosslinked receptor complexes with SDS and immunoprecipitated them with antibodies against SRK₈ or SLG₈ (Fig. 4c). The polyclonal antibody raised against an SRK₈ carboxy-terminal peptide precipitated only the 120K protein; however, the polyclonal antibody against SRK₈ S domain expressed by *E. coli* and the monoclonal antibody against SLG₈ precipitated both 120K and 65K proteins. These results suggest that the 120K protein is SRK₈, and that the 65K protein is either SLG₈ or eSRK₈ (eSRK is a soluble truncated form of SRK produced by alternative splicing and so far found in the S_3 haplotype of *Brassica oleracea*²² and in transgenic tobacco carrying genomic SRK₉ gene of *Brassica rapa*²³). As the monoclonal antibody used in this experiment has a higher affinity for SLG₈ than for SRK₈ (data not shown), the finding that this antibody precipitates the 65K protein more efficiently than the 120K protein suggests that the 65K protein may be SLG₈; however, the presence of eSRK₈ in the 65K band cannot be completely ruled out.

The crosslinking of ^{125}I -labelled S_8 -SP11 to both SRK₈ and SLG₈ indicates that there may be a close association between the latter two

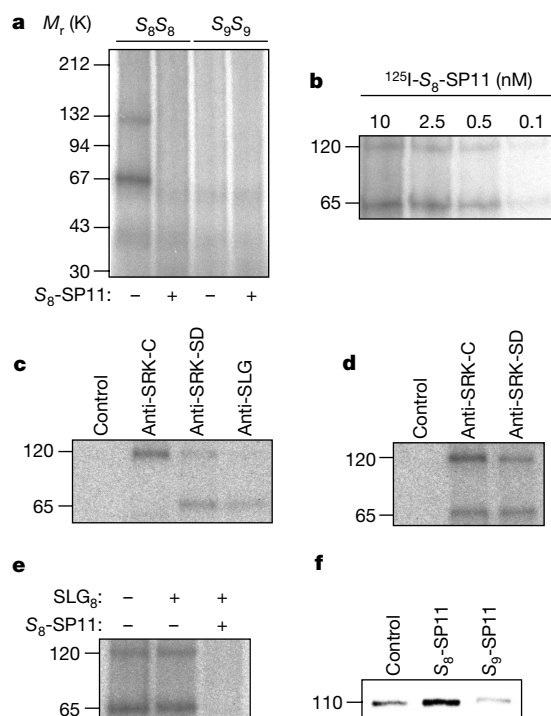


Figure 4 Interaction of S_8 -SP11 with SRK₈–SLG₈ receptor complex. **a**, Chemical crosslinking of ^{125}I -labelled S_8 -SP11 to stigma microsomal membranes of S_8 and S_9 homozygotes. Microsomal membranes were subjected to crosslinking in a solution containing ^{125}I -labelled S_8 -SP11 (10 nM) in the absence or presence of S_8 -SP11 (1 μM) as competitor. **b**, Crosslinking patterns at various concentrations of ^{125}I -labelled S_8 -SP11. **c**, Immunoprecipitation of crosslinked microsomal proteins that had been solubilized with SDS. IgG fraction of pre-immune serum (Control), a polyclonal peptide antibody against the SRK₈ C terminus (Anti-SRK-C), a polyclonal antibody against the SRK₈ S domain (Anti-SRK-SD), and a monoclonal antibody for SLG (Anti-SLG) were used for immunoprecipitation. **d**, Immunoprecipitation of crosslinked microsomal proteins that had been solubilized with Triton X-100. **e**, Crosslinking in the absence or the presence of 100 nM of SLG₈ and S_8 -SP11. **f**, Induction of autophosphorylation of SRK₈ by S_8 -SP11 and S_9 -SP11 (both at 10 nM).

proteins, because BS³ can only crosslink primary amine groups that are less than about 11 Å from each other. A close association between SRK₈ and SLG₈ is further supported by the following observations. When Triton X-100, a non-ionic mild detergent, was used to solubilize crosslinked microsomal membranes instead of SDS, the antibody against the SRK₈ C terminus immunoprecipitated SRK₈ as well as SLG₈ (Fig. 4d). The SLG₈ in the precipitate might have resulted from the contamination of soluble SLG₈ in the microsomal membrane fraction of the S₈ homozygote; however, repeated washing of the microsomal fraction, a procedure known to reduce any contaminating soluble SLG, did not change the crosslinking pattern of ¹²⁵I-labelled S₈-SP11 (data not shown). Thus, the crosslinked SLG₈ is probably not the washable soluble form, but rather the SRK-associated form of SLG—the presence of which has been suggested by velocity sedimentation analysis of stigma extract²⁴.

Furthermore, addition of excess SLG₈ to the microsomal fraction did not affect the crosslinking pattern (Fig. 4e), confirming that the soluble form of SLG₈ has little affinity for S₈-SP11. Together, these results indicate that SRK and SLG form a high-affinity receptor complex for SP11 on the papilla cell membrane. This finding is consistent with our previous observation that SLG enhances the self-incompatibility response⁶ and might explain the finding that co-expression of SLG stabilizes SRK in transgenic plants²⁵.

The involvement of SLG in the formation of the stable high-affinity receptor complex for SP11 might also explain the breakdown of self-incompatibility in two *Brassica* strains that have defects in the structure or expression of SLG²⁵. But this role of SLG may not be essential for all S haplotypes, because two functional S haplotypes of *B. oleracea* have been found to lack a functional SLG²⁶. This suggests that, in some S haplotypes, either SRK alone can transmit the SP11 signal or some other molecules (such as eSRK) can substitute for SLG.

It has been shown that a mixture of PCPs can induce autophosphorylation of SRK; however, the nature of the component(s) in the mixture responsible for the activation was not determined²⁷. Having shown the specific binding of SP11 to its cognate SRK–SLG complex, we further examined whether this interaction results in the activation of SRK. When added to the plasma membrane fraction of S₈ stigmas at a kinetically relevant concentration, S₈-SP11 but not S₉-SP11 induced autophosphorylation of SRK₈ (Fig. 4f), thus showing that SP11 alone can activate SRK in an S-haplotype-specific manner. It should be noted that the previous study mentioned above showed that SRK was kept in an inactivated form by stigmatic thioredoxin, and thus activation of SRK by PCPs could only be observed in the presence of thioredoxin²⁷. However, we have detected SRK activation without the addition of thioredoxin to the plasma membranes. As we used a highly purified plasma membrane fraction that is not likely to contain thioredoxin (see Methods), our results seem to suggest that activation of SRK by SP11 is independent of the negative regulation by thioredoxin.

Plants produce soluble receptor-like proteins that are highly similar in sequence to the extracellular domain of their corresponding receptor kinases¹⁰. These 'truncated' forms are produced either from the same gene by alternative splicing as in the case of eSRK and SRK, or from different genes as in the case of SLG and SRK. In either case, the physiological significance of the soluble form of a receptor kinase is virtually unknown. Our present data provide biochemical evidence that SLG directly participates with SRK in the formation of a high-affinity receptor complex, and that the binding of SP11 ligand to its cognate receptor complex induces the phosphorylation of SRK and self-incompatibility response in the papilla cell. This finding not only has important implications for the biochemical mechanism of self-incompatibility interactions in *Brassica*, but also opens a new direction in the study of receptor–ligand interactions in plants. □

Methods

Isolation of S₈-SP11 protein

Pollen grains were collected from dehiscent anthers of *B. rapa* S₈ homozygotes. The pollen surface materials were dissolved in 1 M AcOH by vortexing, and absorbed onto an SP-Sephadex C-25 (H⁺ form) column. After the column was washed with 1 M AcOH and then with 2 M pyridine, basic proteins were eluted with 2 M pyridine-AcOH (pH 5.0)²⁸ and lyophilized. We immunoprecipitated S₈-SP11 using an S₈-SP11 antibody⁸ that had been pre-conjugated to rProtein A/Sepharose (Amersham Pharmacia) using dimethylpimelidate (Tokyo Chemical). The precipitated S₈-SP11 was eluted with 100 mM glycine-HCl (pH 2.0) and analysed by MALDI-TOF-MS after desalting by Zip-Tip (Millipore).

Synthesis of S₈-SP11 protein

S₈-SP11 was synthesized by Peptide Institute Inc. (Osaka). Briefly, the protected form of S₈-SP11 was assembled from seven building blocks of S₈-SP11 segments that had been synthesized separately by either the solid-phase or the liquid-phase method. After deprotection, S₈-SP11 was reduced in the presence of dithiothreitol (DTT) and purified by reversed-phase HPLC. The reduced form of S₈-SP11 was refolded in the presence of both oxidized and reduced forms of glutathione (1:10). One major oxidized form of S₈-SP11 was purified to homogeneity by reversed-phase HPLC. We determined the positions of the disulphide bonds by analysing proteolytic fragments of S₈-SP11 using a combination of Edman degradation and MALDI-TOF-MS. Details of the chemical synthesis and the primary structural characterization of S₈-SP11 will be presented elsewhere (S.T., H.S., M.I., H.S. and A.L., unpublished data). Non-radioactive iodinated S₈-SP11 (I-S₈-SP11) used in the pollination bioassay was prepared using Iodo-gen (Pierce) and purified by reversed-phase HPLC.

Pollination bioassay

Stigmas were treated with 0.5 µl of an SP11 solution containing 0.05% Tween 20, and then dried in air for 1 h. After cross-pollination, the stigmas were kept at 20 °C for 6 h, and penetration of pollen tubes into each stigma was observed after aniline-blue staining as described²⁹. When more than 10 pollen tubes penetrated a stigma, we judged the pollination to be compatible.

Biosensor measurements

All measurements were performed on a BIAcore 2000 as described¹⁸. SLGs and SLR1 were purified from stigmas as described^{9,18}. For SRK₉ extracellular domain preparation, a GST gene was fused downstream of the region of the SRK₉ gene encoding amino-acid residues 1–428, and the fusion construct was cloned into pBK283 (Nihon Nosan) and co-transfected with BmNPV DNA into BmN4 insect cells. Larvae of *Bombyx mori* L. were infected with amplified recombinant baculoviruses, and the recombinant protein, SRK₉-SD–GST, was affinity purified using a glutathione-sepharose column (Amersham Pharmacia). An SLR1–GST fusion protein was obtained similarly for use as a control. Both GST fusion proteins were immobilized separately to the parallel flow cells of a CM5 sensor chip using a GST kit (Biacore).

Binding experiments

Stigma microsomal fractions were prepared by differential centrifugation in an ice-cold buffer containing 50 mM Tris-HCl pH 7.5, 250 mM sucrose, 10 mM DTT and a protease inhibitor cocktail (Roche). After ultracentrifugation, the microsome pellet was washed twice and stored in a resuspension buffer containing 10 mM Tris-HCl pH 7.5, 250 mM sucrose, 20% glycerol and 0.1% bovine serum albumin (BSA). S₈-SP11 (2 nmol) was iodinated to a specific activity of 6.1 TBq mmol^{−1} with ¹²⁵I using Iodo-gen. Binding reactions were carried out for 90 min on ice in binding buffer containing 10 mM HEPES-KOH pH 7.4, 100 mM NaCl, 100 mM sucrose, 5 mM MgCl₂, 0.1% BSA, and the protease inhibitor cocktail. Specific binding was determined in the presence of a 100-fold excess of S₈-SP11. We separated the bound and free ¹²⁵I-labelled S₈-SP11 by centrifugation.

Crosslinking and immunoprecipitation

The monoclonal antibody against SLG₈ has been described²⁹. Polyclonal antibodies against SRK₈ were produced in rabbits by immunization with a keyhole limpet haemocyanin-conjugated SRK₈ C-terminal peptide (DDESWTNKKYTCSTVIDAR) or an SRK₈ extracellular domain recombinant protein (corresponding to residues 30–435) expressed from *E. coli*, and the antibodies were affinity purified. Crosslinking was performed at room temperature for 30 min with 1 mM BS³ (Pierce). Microsomal membranes were solubilized with 1% SDS or with 1% Triton X-100, and immunoprecipitated by antibodies pre-conjugated to Protein G/Sepharose (Amersham Pharmacia) or rProtein A/Sepharose.

Phosphorylation *in vitro*

We purified plasma membranes from microsomal membranes by the aqueous two-phase partitioning method as described³⁰. Plasma membranes (2 µg) were incubated for 90 min in the presence or absence of 10 nM S₈-SP11 or S₉-SP11. After the membranes were solubilized with 0.5% Nonidet P-40, the SRK complex was recovered using SRK₈ C-terminal peptide antibody and rProtein A/Sepharose. We incubated the immunoprecipitated SRK complex for 5 min on ice in phosphorylation buffer 20 mM HEPES-KOH

pH 7.4, 100 mM NaCl, 0.1% Nonidet P-40, 10% glycerol, 5 mM MgCl₂, 5 mM MnCl₂, 10 mM NaF, 0.1 mM Na₂VO₄, 5 µM microcystin LR, 1 µM okadaic acid, 1 mM DTT, the protease inhibitor cocktail and 0.33 µCi µl⁻¹ [γ-³²P]ATP (Amersham).

Received 20 April; accepted 2 August 2001.

- McCubbin, A. G. & Kao, T.-h. Molecular recognition and response in pollen and pistil interactions. *Annu. Rev. Cell Dev. Biol.* **16**, 333–364 (2000).
- Stein, J. C., Howlett, B., Boyes, D. C., Nasrallah, M. E. & Nasrallah, J. B. Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. *Proc. Natl Acad. Sci. USA* **88**, 8816–8820 (1991).
- Suzuki, G. *et al.* Genomic organization of the *S* locus: identification and characterization of genes in *SLG/SRK* region of *S₀* haplotype of *Brassica campestris* (syn. *rapa*). *Genetics* **153**, 391–400 (1999).
- Schopfer, C. R., Nasrallah, M. E. & Nasrallah, J. B. The male determinant of self-incompatibility in *Brassica*. *Science* **286**, 1697–1700 (1999).
- Nasrallah, J. B., Kao, T.-H., Chen, C.-H., Goldberg, M. L. & Nasrallah, M. E. Amino-acid sequence of glycoproteins encoded by three alleles of the *S* locus of *Brassica oleracea*. *Nature* **326**, 617–619 (1987).
- Takasaki, T. *et al.* The *S* receptor kinase determines self-incompatibility in *Brassica stigma*. *Nature* **403**, 913–916 (2000).
- Takayama, S. *et al.* The pollen determinant of self-incompatibility in *Brassica campestris*. *Proc. Natl Acad. Sci. USA* **97**, 1920–1925 (2000).
- Shiba, H. *et al.* A pollen coat protein, SP11/SCR, determines the pollen *S*-specificity in the self-incompatibility of *Brassica* species. *Plant Physiol.* **125**, 2095–2103 (2001).
- Takayama, S. *et al.* Sequences of *S*-glycoproteins, products of the *Brassica campestris* self-incompatibility locus. *Nature* **326**, 102–104 (1987).
- Torii, K. U. Receptor kinase activation and signal transduction in plants: an emerging picture. *Curr. Opin. Plant Biol.* **3**, 361–367 (2000).
- Dickinson, H. G. Pollen stigma interactions: so near yet so far. *Trends Genet.* **16**, 373–376 (2000).
- Brugière, N., Cui, Y. & Rothstein, S. J. Molecular mechanisms of self-recognition in *Brassica* self-incompatibility. *Trends Plant Sci.* **5**, 432–438 (2000).
- Broekaert, W. F., Terras, F. R. G., Cammue, B. P. A. & Osborn, R. W. Plant defensins: novel antimicrobial peptides as components of the host defense system. *Plant Physiol.* **108**, 1353–1358 (1995).
- Bruix, M. *et al.* Solution structure of γ1-H and γ1-P thionins from barley and wheat endosperm determined by ¹H-NMR: A structural motif common to toxic arthropod proteins. *Biochemistry* **32**, 715–724 (1993).
- Watanabe, M. *et al.* Highly divergent sequences of the pollen self-incompatibility (*S*) gene in class-I *S* haplotypes of *Brassica campestris* (syn. *rapa*) L. *FEBS Lett.* **473**, 139–144 (2000).
- Stephenson, A. G. *et al.* The male determinant of self-incompatibility in *Brassica oleracea* is located in the pollen coating. *Plant J.* **12**, 1351–1359 (1997).
- Doughty, J. *et al.* PCP-A1, a defensin-like *Brassica* pollen coat protein that binds the *S* locus glycoprotein, is the product of gametophytic gene expression. *Plant Cell* **10**, 1333–1347 (1998).
- Takayama, S. *et al.* Isolation and characterization of pollen coat proteins of *Brassica campestris* that interact with *S* locus-related glycoprotein 1 involved in pollen-stigma adhesion. *Proc. Natl Acad. Sci. USA* **97**, 3765–3770 (2000).
- Trutochaud, A. E., Jeong, S. & Clark, S. E. CLAVATA3, a multimeric ligand for the CLAVATA1 receptor-kinase. *Science* **289**, 613–617 (2000).
- Gómez-Gómez, L., Bauer, Z. & Boller, T. Both the extracellular leucine-rich repeat domain and the kinase activity of FLS2 are required for flagellin binding and signaling in *Arabidopsis*. *Plant Cell* **13**, 1155–1163 (2001).
- Wang, Z.-Y., Seto, H., Fujioka, S., Yoshida, S. & Chory, J. BRI1 is a critical component of a plasma-membrane receptor for plant steroids. *Nature* **410**, 380–383 (2001).
- Giranton, J.-L., Ariza, M. J., Dumas, C., Cock, J. M. & Gaude, T. The *S* locus receptor kinase gene encodes a soluble glycoprotein corresponding to the SRK extracellular domain in *Brassica oleracea*. *Plant J.* **8**, 827–834 (1995).
- Suzuki, G., Watanabe, M., Toriyama, K., Isogai, A. & Hinata, K. Expression of *SLG₀* and *SRK₀* genes in transgenic tobacco. *Plant Cell Physiol.* **37**, 866–869 (1996).
- Giranton, J.-L., Dumas, C., Cock, J. M. & Gaude, T. The integral membrane *S*-locus receptor kinase of *Brassica* has serine/threonine kinase activity in a membranous environment and spontaneously forms oligomers in planta. *Proc. Natl Acad. Sci. USA* **97**, 3759–3764 (2000).
- Dixit, R., Nasrallah, M. E. & Nasrallah, J. B. Post-transcriptional maturation of the *S* receptor kinase of *Brassica* correlates with co-expression of the *S*-locus glycoprotein in the stigmas of two *Brassica* strains and in transgenic tobacco plants. *Plant Physiol.* **124**, 297–311 (2000).
- Suzuki, T., Kusaba, M., Matsushita, M., Okazaki, K. & Nishio, T. Characterization of *Brassica S*-haplotypes lacking *S*-locus glycoprotein. *FEBS Lett.* **482**, 102–108 (2000).
- Cabrillac, D., Cock, J. M., Dumas, C. & Gaude, T. The *S*-locus receptor kinase is inhibited by thioredoxins and activated by pollen coat proteins. *Nature* **410**, 220–223 (2001).
- Kojima, M. *et al.* Ghrelin is a growth-hormone-releasing acylated peptide from stomach. *Nature* **402**, 656–660 (1999).
- Shiba, H. *et al.* Alteration of the self-incompatibility phenotype in *Brassica* by transformation of the antisense *SLG* gene. *Biosci. Biotechnol. Biochem.* **64**, 1016–1024 (2000).
- Larsson, C., Widell, S. & Kjellbom, P. Preparation of high-purity plasma membranes. *Methods Enzymol.* **148**, 558–568 (1987).

Acknowledgements

We thank T.-H. Kao for critically reading and editing the manuscript; and H. Sugita, T. Ueda, K. Iwasaki and H. Sato for technical assistance. This work was supported in part by grants-in-aid from the Ministry of Education, Culture, Sports, Science and Technology, and grants from The Japan Society for the Promotion of Science.

Correspondence and requests for material should be addressed to A.I. (e-mail: isogai@bs.aist-nara.ac.jp).

Quality control of mRNA 3'-end processing is linked to the nuclear exosome

Patricia Hilleren*, Terri McCarthy†, Michael Rosbash†, Roy Parker* & Torben Heick Jensen†‡

* Howard Hughes Medical Institute, Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona 85721, USA

† Howard Hughes Medical Institute, Department of Biology, Brandeis University, Waltham, Massachusetts 02454, USA

An emerging theme in messenger RNA metabolism is the coupling of nuclear pre-mRNA processing events, which contributes to mRNA quality control¹. Most eukaryotic mRNAs acquire a poly(A) tail during 3'-end processing within the nucleus, and this is coupled to efficient export of mRNAs to the cytoplasm^{2,3}. In the yeast *Saccharomyces cerevisiae*, a common consequence of defective nuclear export of mRNA is the hyperadenylation of nascent transcripts^{4,5}, which are sequestered at or near their sites of transcription⁵. This implies that polyadenylation and nuclear export are coupled in a step that involves the release of mRNA from transcription site foci. Here we demonstrate that transcripts which fail to acquire a poly(A) tail are also retained at or near transcription sites. Surprisingly, this retention mechanism requires the protein Rrp6p and the nuclear exosome, a large complex of exonucleolytic enzymes^{6,7}. In exosome mutants, hypo- as well as hyperadenylated mRNAs are released and translated. These observations suggest that the exosome contributes to a checkpoint that monitors proper 3'-end formation of mRNA.

To test whether unadenylated transcripts would also be retained within the nucleus, we used a strain carrying a lesion in the poly(A) polymerase gene, *pap1-1* (ref. 8), to generate poly(A)⁻ mRNA, which could then be localized by fluorescence *in situ* hybridization (FISH). We chose to examine *SSA4* RNA, which is synthesized *de novo* when the strain is shifted to the restrictive temperature. The poly(A)⁻ *SSA4* transcripts accumulated in discrete intranuclear foci (Fig. 1a), indistinguishable from the localization of hyperadenylated *SSA4* transcripts⁵. Comparable results were obtained with probes directed against *HSP104*, another heat-induced transcript (data not shown). We conclude that both hypo- and hyperadenylation causes transcript retention at or near transcription sites.

The sequestration of both poly(A)⁻ and hyperadenylated mRNAs in transcription site foci indicated that a system exists in yeast to both monitor the quality of 3'-end formation and inhibit the release of aberrant transcripts. Strains defective in such a quality-control system should fail to retain aberrant RNAs at these foci. Because defects in the nonessential nuclear exonuclease Rrp6p partially rescue the temperature sensitivity of *pap1-1* (ref. 9), we investigated the effect of an *RRP6* deletion on transcript localization. The sequestration of *SSA4* poly(A)⁻ RNA was markedly absent in the *pap1-1 rrp6Δ* strain (Fig. 1a). It is important to note that this change in *SSA4* localization occurred without any dramatic change in mRNA structure or abundance, as shown by northern analysis (Fig. 1b). Moreover, the loss of Rrp6p partially rescued the block to heat-shock protein synthesis caused by the *pap1-1* lesion (Fig. 1c, compare lanes 6 and 8). This indicates that the absence of Rrp6p allows both release and function of the poly(A)⁻ mRNA.

To assay another transcript, we examined *PGK1pG* mRNA. Because of its inducible promoter, we could use a transcriptional pulse-chase protocol to produce a synchronous pool of mRNA¹⁰.

‡ Present address: Department of Molecular and Structural Biology, University of Aarhus, DK-8000 Aarhus C, Denmark.

A poly(G) structure engineered into the 3' untranslated region of *PGK1pG* permits detection of decay intermediates produced by mRNA turnover. Similar to *SSA4* RNA (Fig. 1b), the predominant mRNA produced in the *pap1-1* and *pap1-1 rrp6Δ* strains was poly(A)⁻ (Fig. 1d), although some poly(A)⁺ species of *PGK1pG* were produced. This probably reflects residual polyadenylation by the mutant polymerase under restrictive conditions. Although the poly(A)⁻ species was surprisingly stable in the *pap1-1* strain, the poly(A)⁻ transcripts were destabilized in the absence of Rrp6p (Fig. 1d), and a decay intermediate accumulated. These observations demonstrate that the fate of the poly(A)⁻ mRNA was changed in the *rrp6Δ* strain. Importantly, because the *PGK1pG* mRNA is destabilized in the absence of Rrp6p, these observations are inconsistent with Rrp6p functioning solely to degrade nuclear poly(A)⁻ mRNAs. We conclude that sequestration of poly(A)⁻ mRNA in transcription site foci requires the activity of Rrp6p.

To investigate whether Rrp6p is also required for retention of hyperadenylated transcripts, we used the *rat7-1* and *rip1Δ* export mutant strains. In these strains, the *SSA4* and *HSP104* transcripts are hyperadenylated and retained at or near transcription sites (Fig. 2a and ref. 5). In the *rat7-1 rrp6Δ* and the *rip1Δ rrp6Δ* strains, both the *HSP104* and *SSA4* mRNAs were still nuclear at 42 °C but much more diffuse than the discrete nuclear foci observed in the single export mutant strains (Fig. 2a and data not shown). Inter-

estingly, in the *rat7-1 rrp6Δ* and *rip1Δ rrp6Δ* strains, heat shock mRNAs accumulated in a subregion of the nucleus, forming a cap on top of the DAPI stain (Fig. 2a), suggesting these transcripts might be accumulating in a subnuclear location. This suggests further that, as for transcripts produced in *pap1-1*, transcripts in *rat7-1* and *rip1Δ* are also released from transcription site foci in the absence of Rrp6p. In contrast to *pap1-1* cells, however, the mRNA export defects are not relieved by deletion of *RRP6*, which causes at least some of these transcripts to remain within the nucleus. Also in contrast to the *pap1-1* data, a clear increase in *SSA4* transcript levels was observed in the *rat7-1 rrp6Δ* and *rip1Δ rrp6Δ* strains compared with the single mutants (Fig. 2b), suggesting that Rrp6p may contribute to the rapid degradation of some hyperadenylated transcripts. A less likely possibility is that the absence of Rrp6p leads to an increase in transcription rate.

Further support for transcript release in *rrp6Δ* strains is the observation that heat-shock protein synthesis is partially restored in the *rat7-1 rrp6Δ* strain (Fig. 2c, lane 8). A low level of heat-shock protein expression is observed in the *rat7-1* cells (Fig. 2c, lane 6), supporting the previous claim that mRNA export is leaky in this strain⁴. In contrast, heat-shock protein synthesis is not restored in the *rip1Δ rrp6Δ* strain, consistent with a tighter mRNA export block in the *rip1Δ* background (Fig. 2c and data not shown).

Rrp6p is a member of the nuclear exosome, an exonucleolytic

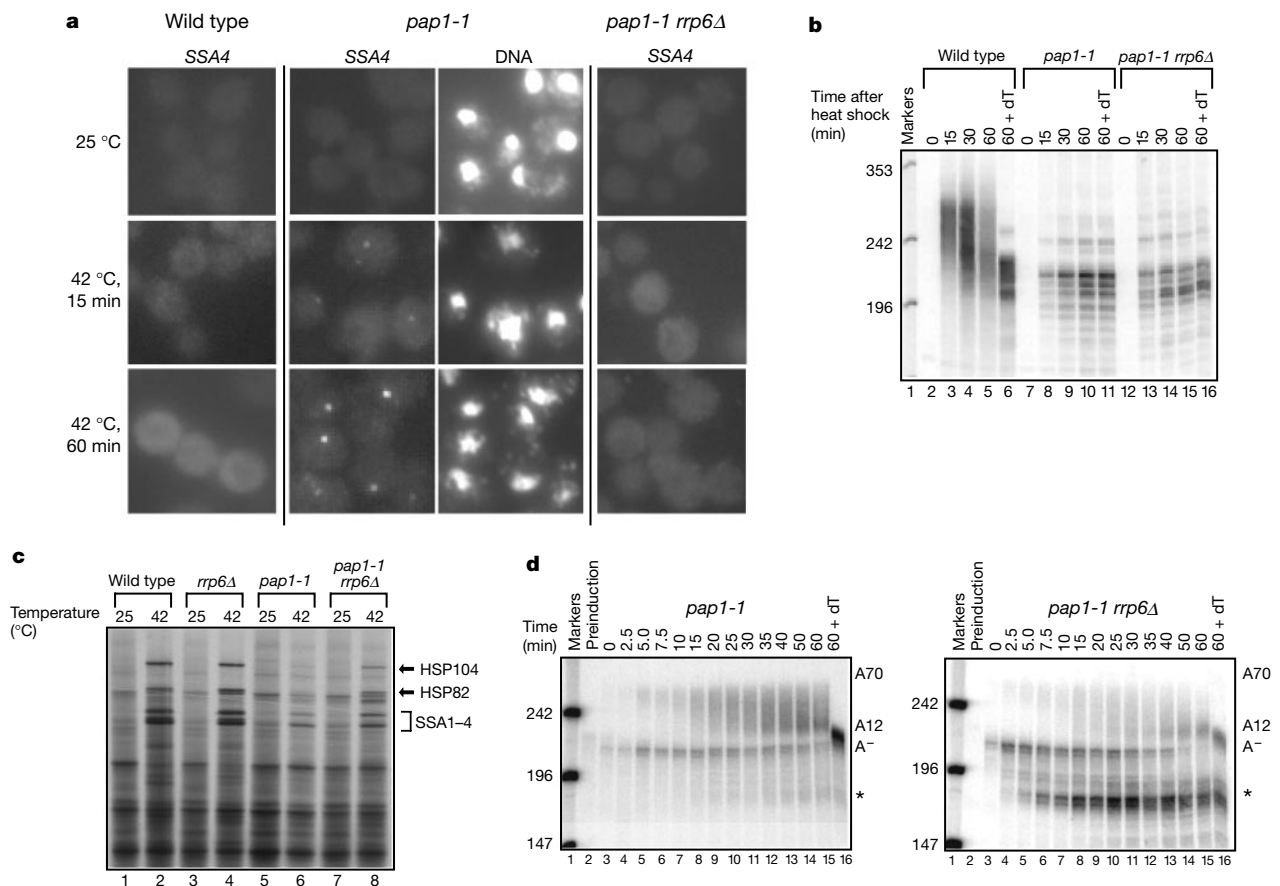


Figure 1 *SSA4* poly(A)⁻ mRNA accumulates in an intranuclear focus in an Rrp6p-dependent manner. **a**, *SSA4* mRNA FISH analysis of wild type, *pap1-1* and *pap1-1 rrp6Δ* cultures grown at 25 °C or temperature shifted to 42 °C for 15 min or 60 min, as indicated. DNA was stained with DAPI. **b**, Northern analysis of *SSA4* mRNA isolated from cultures collected at the indicated time points after a 42 °C heat shock. dT indicates that the RNA sample was treated with oligo(dT) and RNaseH before gel electrophoresis. The multiple poly(A)⁻ species produced in the *pap1-1* strains is probably due to promiscuous 3'-end cleavage^{17,18}. The sizes of radiolabelled molecular mass markers in nucleotides are indicated to the left of the panel. **c**, Heat-shock protein synthesis in wild-type, *rrp6Δ*,

pap1-1 and *pap1-1 rrp6Δ* strains grown at 25 °C or temperature shifted to 42 °C for 15 min. Migration of the major heat-shock proteins is shown to the right of the panel. **d**, Transcriptional pulse-chase of *PGK1pG* in *pap1-1* and *pap1-1 rrp6Δ* strains. Cultures were grown in 2% raffinose. Transcription was induced by the addition of galactose to a final concentration of 2%. After 5 min, transcription was repressed by the addition of glucose to a final concentration of 4%. The position of the major poly(A)⁻ species is indicated to the right of each panel. The position of a decay intermediate is indicated by an asterisk to the right of each panel.

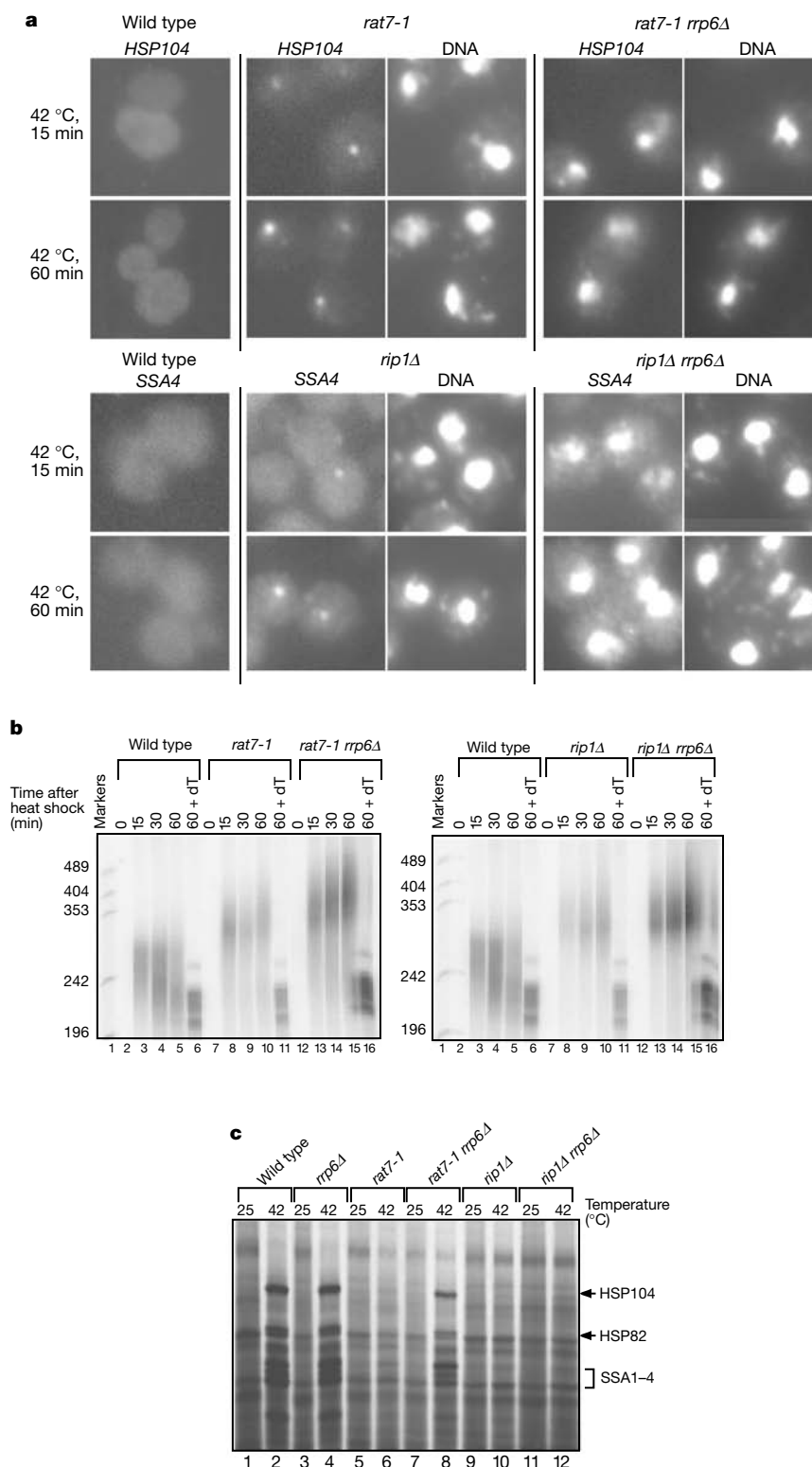


Figure 2 Rrp6p mediates transcription site foci retention of *HSP104* and *SSA4* mRNAs in mRNA export mutants. **a**, *HSP104* mRNA FISH analysis on the *rat7-1* and *rat7-1 rrp6Δ* strains, and *SSA4* mRNA FISH analysis on *rip1Δ* and *rip1Δ rrp6Δ* strains are shown. Experiments were performed as described in Fig. 1a. **b**, Northern analysis of *SSA4* mRNA

was performed as described in Fig. 1b. dT indicates treatment with oligo(dT) and RNaseH. The sizes of radiolabelled molecular mass markers in nucleotides are indicated to the left of the panel. **c**, Heat-shock protein synthesis in the indicated strains was analysed as described in Fig. 1c.

complex that contributes to a variety of nuclear RNA processing and degradative reactions^{6,7,11–13}. To examine the effects of other nuclear exosome components, we analysed the fate of hyperadenylated transcripts in temperature-sensitive mutants of the essential RRP4 and MTR4 genes. At 42 °C, *HSP104* localization in the *rat7-1 rrp4-1*

and *rat7-1 mtr4-1* double-mutant strains parallels that seen in the *rat7-1 rrp6Δ* strain (Fig. 3a). The similar results with three exosome components suggest that retention of aberrant mRNAs is a function of the nuclear exosome. Consistent with this notion, the *rat7-1 rrp6Δ*, *rat7-1 rrp4-1* and *rat7-1 mtr4-1* double-mutant strains all

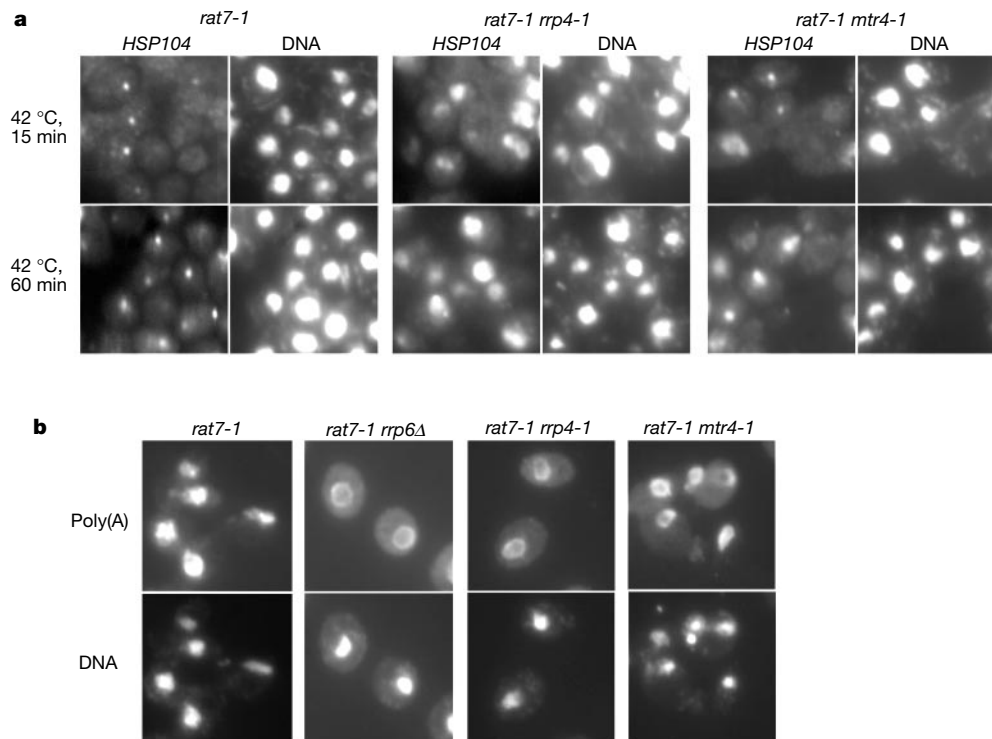


Figure 3 Lesions in the nuclear exosome alter the distribution of mRNA in the *rat7-1* background. **a**, *HSP104* mRNA FISH analysis on the *rat7-1*, *rat7-1 rrp4-1* and *rat7-1 mtr4-1* strains temperature shifted to 42 °C for 15 or 60 min are shown. **b**, Oligo(dT) FISH

analyses of *rat7-1*, *rat7-1 rrp6Δ*, *rat7-1 rrp4-1* and *rat7-1 mtr4-1* cultures temperature shifted to 37 °C for 30 min are shown.

display an altered distribution of the total poly(A)⁺ signal. As previously reported and shown in Fig. 3b, *rat7-1* single-mutant cells exhibit a strong intranuclear granular poly(A)⁺ staining after a shift to 37 °C (ref. 14). In contrast, all three double-mutant strains show peri-nuclear poly(A)⁺ staining. This localization probably reflects release of the transcripts from transcription site foci and a persistent block in mRNA export at the nuclear periphery, directly due to the *rat7-1* mutation. These comparisons further strengthen the conclusion that aberrant transcripts are blocked near transcription sites in *rat7-1* cells and released when the nuclear exosome is compromised.

Our results indicate that mRNAs with aberrant 3' ends are generally retained at or near their sites of transcription. Moreover, we propose that the nuclear exosome contributes to a quality-control system that monitors correct 3'-end formation before RNA release from these sites. It is interesting to note that Rrp6p interacts genetically and physically with Pap1p as well as with the nucleocytoplasmic shuttling protein, Npl3p (refs 9, 15), suggesting that some of the links between the exosome, 3'-end formation and nuclear export are direct. Given that the exosome functions in nuclear mRNA turnover¹³, an intriguing possibility is that the exosome participates not only in the destruction of aberrant mRNAs but also in their detection and retention at or near transcription sites. A remaining challenge is to understand the physical link between incorrectly adenylated mRNA and their retention, that is, why this mRNA is not free to diffuse away from the gene locus. □

Methods

Yeast methods

Strains used in this analysis are listed in the Supplementary Information as Table 1. Double mutants were generated by standard genetic crosses.

In situ hybridization analysis

FISH of *SSA4* and *HSP104* mRNA was performed using Cy3-labelled probes, and DNA was stained with DAPI as previously described⁵. Poly(A)⁺ RNA was localized using a Cy3-labelled oligo(dT)₇₀ probe¹⁶.

RNA and protein analysis

Transcriptional pulse-chase experiments, RNaseH treatment and northern analysis of *PGK1pG* were as previously described^{4,10}. RNaseH treatment and northern analysis of *SSA4*, as well as analysis of heat-shock protein synthesis, were performed as described⁵.

Received 20 April; accepted 9 August 2001.

- Maquat, L. E. & Carmichael, G. G. Quality control of mRNA function. *Cell* **104**, 173–176 (2001).
- Eckner, R., Ellmeier, W. & Birnstiel, M. L. Mature mRNA 3' end formation stimulates RNA export from the nucleus. *EMBO J.* **10**, 3513–3522 (1991).
- Huang, Y. & Carmichael, G. C. Role of polyadenylation in nucleocytoplasmic transport of mRNA. *Mol. Cell. Biol.* **16**, 1534–1542 (1996).
- Hillgren, P. & Parker, R. Defects in mRNA export factors Rat7p, Gle1p, Mex67p and Rat8p cause hyperadenylation during 3' end formation of nascent transcripts. *RNA* **7**, 753–764 (2001).
- Jensen, T. H., Patricio, K., McCarthy, T. & Rosbash, M. A block to mRNA nuclear export in *Saccharomyces cerevisiae* leads to hyperadenylation of transcripts that accumulate at the site of transcription. *Mol. Cell* **7**, 887–898 (2001).
- Briggs, M. W., Burkard, K. T. & Butler, J. S. Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. *J. Biol. Chem.* **273**, 13255–13263 (1998).
- Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollervey, D. The exosome: A conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonucleases. *Cell* **91**, 457–466 (1997).
- Patel, D. & Butler, J. S. Conditional defect in mRNA 3' end processing caused by a mutation in the gene for poly(A) polymerase. *Mol. Cell. Biol.* **12**, 3297–3304 (1992).
- Burkard, K. T. & Butler, J. S. A nuclear 3'→5' exonuclease involved in mRNA degradation interacts with poly(A) polymerase and the hnRNA protein Npl3p. *Mol. Cell. Biol.* **20**, 604–616 (2000).
- Decker, C. J. & Parker, R. A turnover pathway for both stable and unstable mRNAs in yeast: Evidence for a requirement for deadenylation. *Genes Dev.* **7**, 1632–1643 (1993).
- Allmang, C. *et al.* Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J.* **18**, 5399–5410 (1999).
- van Hoof, A., Lennertz, P. & Parker, R. Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol. Cell. Biol.* **20**, 441–452 (2000).
- Bousquet-Antonelli, C., Presutti, C. & Tollervey, D. Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell* **102**, 765–775 (2000).
- Gorsch, L. C., Dockendorff, T. C. & Cole, C. N. A conditional allele of the novel repeat-containing yeast nucleoporin RAT7/NUP159 causes both rapid cessation of mRNA export and reversible clustering of nuclear pore complexes. *J. Cell Biol.* **129**, 939–955 (1995).
- Lee, M. S., Henry, M. & Silver, P. A. A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. *Genes Dev.* **10**, 1233–1246 (1996).
- Vainberg, I. E., Dower, K. & Rosbash, M. Nuclear export of heat shock and non-heat-shock mRNA occurs via similar pathways. *Mol. Cell. Biol.* **20**, 3996–4005 (2000).
- Mandart, E. & Parker, R. Effects of mutations in the *Saccharomyces cerevisiae* RNA14, RNA15, and PAPI genes on polyadenylation *in vivo*. *Mol. Cell. Biol.* **15**, 6979–6986 (1995).
- Amrani, N., Dufour, M. E., Bonneaud, N. & Lacroute, F. Mutations in STS1 suppress the defect in 3' mRNA processing caused by the *rna15-2* mutation in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **252**, 552–562 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank I. Mattaj for critically reading the manuscript. This work was supported by the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to P.H. (e-mail: hilleren@u.arizona.edu) or T.H.J. (e-mail: thj@mbio.aau.dk).

Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain

Neil M. Ferguson, Christl A. Donnelly & Roy M. Anderson

Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College Science, Technology and Medicine, St Mary's Campus, Norfolk Place, London W2 1PG, UK

The foot and mouth disease (FMD) epidemic in British livestock remains an ongoing cause for concern, with new cases still arising in previously unaffected areas. Epidemiological analyses^{1–3} have been vital in delivering scientific advice to government on effective control measures. Using disease, culling and census data on all livestock farms in Great Britain, we analysed the risk factors determining the spatiotemporal evolution of the epidemic and of the impact of control policies on FMD incidence. Here we show that the species mix, animal numbers and the number of distinct land parcels in a farm are central to explaining regional variation in transmission intensity. We use the parameter estimates thus obtained in a dynamical model of disease spread to show that extended culling programmes were essential for controlling the epidemic to the extent achieved, but demonstrate that the epidemic could have been substantially reduced in scale had the most efficient control measures been rigorously applied earlier.

The FMD epidemic (Fig. 1a) peaked in early April 2001 after two months of rapid spread throughout Great Britain following introduction of the O Pan Asian strain of the virus (by an as-yet-undetermined route) in early February. The subsequent decay of the epidemic was initially rapid, but then slowed because of significant new outbreaks in previously little-affected regions (notably North Yorkshire and Lancashire) outside the three major foci in the north (Cumbria, Dumfries and Galloway, and Northumberland), the southwest (Devon and Somerset) and Welsh borders (Herefordshire, Worcestershire and Powys). Emergence of cases in previously unaffected areas and re-emergence in areas believed to have been cleared of infection continue to make elimination of the epidemic difficult.

The policy for control has evolved over the course of the epidemic, and the efficiency and necessity of individual measures have been the topic of ongoing debate^{3–6}. On 23 February, animal movement restrictions were imposed and public access to affected areas was restricted. This was coupled with slaughter of animals on infected premises (IPs) and farms identified as having had dangerous contacts (DCs) with an IP, together with the introduction of biosecurity measures (such as, the use of disinfectant on clothing, boots and farm vehicles). These measures slowed spread but were insufficient to reverse it, largely owing to long delays between the report of possible FMD, confirmation of infection, and culling of animals on the affected farm^{1,2}. With assistance from the army in the logistics of culling large numbers of animals, in late March the

policy was strengthened to aim for culling of animals on IPs and contiguous premises (CPs) within 24 and 48 h, respectively, without waiting for laboratory confirmation of infection. These policies were informed, in part, by model-based epidemiological analyses of

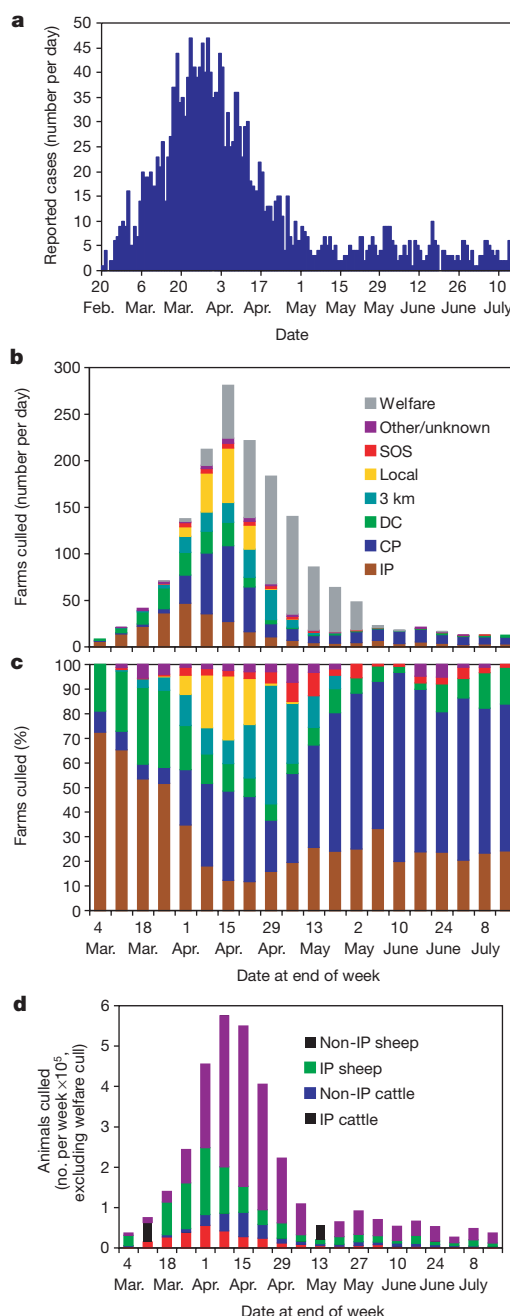


Figure 1 FMD incidence and culling through time. **a**, Time series of FMD cases in the 2001 epidemic, by date of report. **b**, Weekly number of farms culled by control policies: infected premises (IP); contiguous premises (CP); farms identified through tracing as having animals that had dangerous contact (DC) with animals from an IP; farms within 3 km of an IP (3 km); voluntary local sheep cull on farms in heavily affected areas not covered by the 3-km cull (local); farms slaughtered on suspicion (SOS) of infection but not subsequently confirmed by culture or enzyme-linked immunosorbent assay; and other/unknown. Also shown is the weekly number of farms culled to relieve the animal welfare problems resulting from FMD-related restrictions on animal movements (welfare). The 3-km and local culls were largely confined to Cumbria, Dumfries and Galloway. Farms are included only once, so where farms underwent multiple separate culls, the date of the largest cull is used. **c**, Weekly percentage of non-IP farms culled, by policy. **d**, Weekly number of sheep and cattle culled on IP and non-IP farms to control FMD. See Methods and Supplementary Information for details of data sources.

viral spread and the potential impact of various combinations of interventions on the total size of the epidemic and the number of animals culled^{1,7}. Additional culling of sheep on farms within 3 km of an IP also took place in heavily affected areas, together with significant voluntary slaughter of sheep in the same regions (Cumbria, Dumfries and Galloway) from farms outside 3 km. After opposition from some farming groups, vaccination has not been used during the epidemic. Overall, the policies adopted resulted in the culling of 3.49 million animals on some 7,724 farms by 16 July 2001 (and 3.88 million animals by 19 September).

Levels of culling under different schemes varied considerably over time (Fig. 1b–d). In particular, the ratio of non-IP to IP culling (Fig. 1c) was initially low, then peaked in early April as the epidemic began to decline and culling backlogs were cleared, before settling at a somewhat lower level. The ratio of non-IP to IP culling was consistently much higher for sheep than for cattle (Fig. 1d). Overall, up to mid-May, as many animals were culled under the 3-km and local culls as under the CP culling scheme. In late April, the overall level of CP culling dropped (Fig. 1c), coincident with veterinarians being given discretion^{8,9} in determining whether to cull cattle on CPs. However, following the resurgence in case numbers in early May, this relaxation was largely reversed and CP culling increased to the highest level (per FMD case) yet achieved. However, in no week of this epidemic have more than half of CPs been culled within the specified target of 48 h from report of a new case.

The effect of culling on disease transmission cannot be inferred from raw culling rates alone, because the species culled and the spatiotemporal locality of culled farms to recent IPs largely determine the impact of policies. Furthermore, the impact of any particular level of culling depends on the overall intensity of transmission in a region, itself a nonlinear function of epidemiological and demographic parameters. Three related epidemiological parameters are central: (1) the intervention-adjusted basic reproduction number, R_0^I , which reflects the impact of movement restrictions, biosecurity measures and speed of IP culling on transmission intensity (and hence is smaller than the pristine basic reproduction number, R_0); (2) the effective (or case) reproduction number R , which also takes into account loss of susceptible farms through infection and culling; and (3) the average daily transmission coefficient per infectious farm, β . Variation in the speed of IP culling is reflected in R_0^I and R (although not in β), whereas the effectiveness of movement restrictions and biosecurity measures is reflected in all three quantities. Disease elimination requires $R < 1$ in all affected regions¹⁰. Furthermore, $R_0^I < 1$ is necessary to prevent the disease from potentially spreading to previously unaffected areas.

To estimate these parameters and their major determinants, we undertook an integrated analysis of epidemiological and demographic databases (see Methods). The analysis framework used a spatially explicit per-farm hazard model and was formulated to

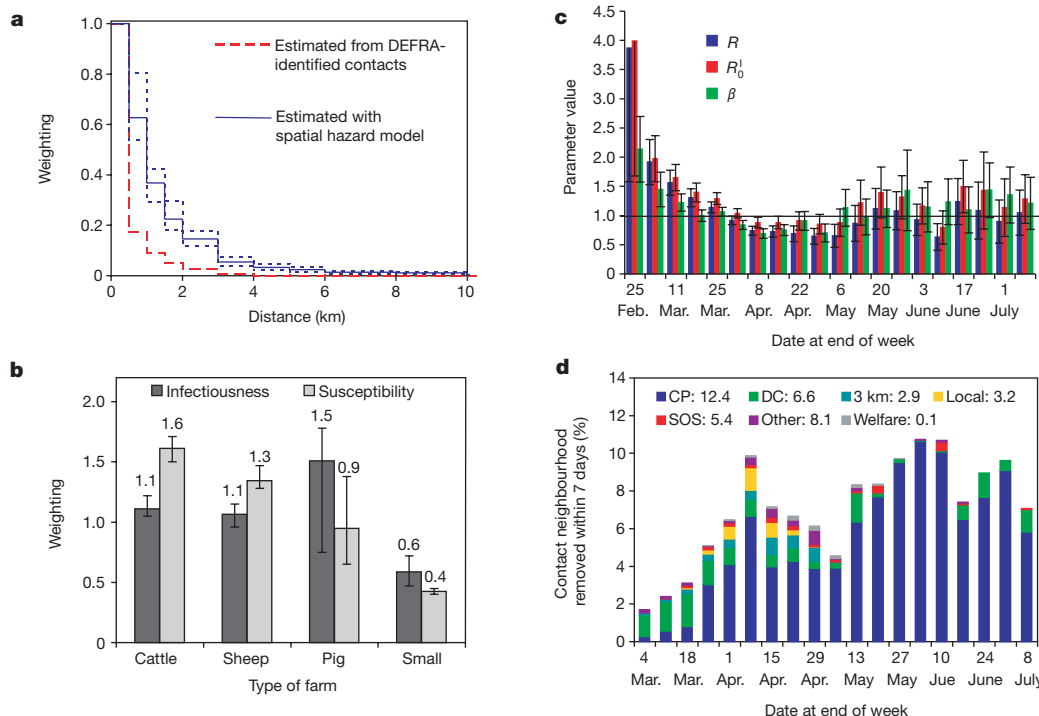


Figure 2 Estimates of transmission parameters. **a**, The spatial transmission kernel (= multiplicative relative risk of transmission as a function of distance from an IP) previously estimated from the distribution of distances between DEFRA-identified infectious contacts¹, and estimated with the spatial hazard model described in the text, with likelihood-based 95% confidence bounds (blue dotted lines). **b**, Relative infectiousness and susceptibility by farm type. Farms were stratified (cattle, sheep, pig and small) on the basis of species numbers, n_c , n_s and n_p , respectively (where n_s represents sheep numbers divided by three to adjust for larger typical stocking numbers), with farms with $n_c + n_s + n_p < 100$ being 'small' and all others being stratified on the basis of the most prevalent species (comparing n_c , n_s and n_p). This stratification was chosen to give (with the exception of pig farms) similar numbers of cases in each category, and because the data often did not support evaluation of the absence of particular species on a farm (zero values in the census or DCS database were often not reproduced in the

other data set). Alternative stratifications gave similar results. **c**, Weekly estimates of R_0^I , R and β with 95% confidence intervals. R was significantly below the threshold for spread (1) throughout April (as daily incidence declined markedly), but rose to about 1 in early May (reflecting roughly constant observed incidence). **d**, Estimated percentage of the contact neighbourhood of an IP (the farms that might be infected by the IP, weighted by susceptibility and infectiousness factors) removed within 7 days of the slaughter of its animals (values are about halved if culls within 48 h of IP slaughter are examined). Numbers after policy names in the chart labels represent policy efficiency (percentage of contact neighbourhoods of all IPs per 1,000 farms slaughtered), which is greatest for CP culling. These estimates are based solely on farm type, fragmentation and location, and so may underestimate the impact of DC culling where a risk of transmission between two farms is suspected from detailed epidemiological investigation.

allow simultaneous estimation of the spatial transmission kernel, infectiousness (which is dependent on farm type, specified by the species mix and number of animals on a farm), susceptibility (which is dependent on both farm type and farmland fragmentation) and time-varying transmission rates (where variation could arise from changes in weather, levels of animal movements and adherence to biosecurity measures).

Farm susceptibility and infectiousness both varied significantly, with smaller farms being substantially less infectious and less susceptible than larger ones (Fig. 2b). Stratifying farms by the most prevalent species revealed a trend for 'cattle' farms to be most susceptible, and 'pig' farms to be least susceptible. The newly estimated spatial kernel differed significantly from that previously derived from the infectious contacts identified by the Department of Environment, Food and Rural Affairs (DEFRA)¹, with considerably more long-distance transmission events (Fig. 2a) being predicted. This implies significant biases in the DEFRA contact-tracing process, with closer contacts being more easily identified. No significant regional or temporal differences in the spatial transmission kernel were found, although statistical power was limited in areas outside Cumbria by lower case incidence. The median distance of the newly estimated kernel is about 4 km, suggesting that most transmission probably occurred through the movement of animals, personnel or vehicles, rather than through animal contact or wind-borne spread.

Except during April, estimates of R_0^1 were consistently greater than or equal to one (Fig. 2c), indicating that IP culling (even within 24 h) and movement restrictions were insufficient to control spread, and that infections seeded into new areas (not yet subject to either IP or non-IP culling) would typically cause local spread. The impact of the depletion of farms in the neighbourhoods of IPs (due to both IP and non-IP culling) can be assessed by comparing estimates of R_0^1 with those of R (Fig. 2c). Reflecting the trends in the incidence of cases, R was more than one until late March, less than one throughout April but rose again in May (reflecting the outbreaks in Settle and Clitheroe). Our analyses suggest that the latter increase, observed in both R_0^1 and R , was not due to any substantial relaxation in the speed of IP culling (Fig. 2c), but reflected an underlying increase in β . This dropped in late February, reflecting the impact of movement restrictions, and then fortuitously fell again still further in early April (which contributed substantially to the overall decline in the epidemic), before rising in early May.

Warmer weather may decrease the viability of virus¹¹, and may explain some of the drop in April, but the May increase can be attributed easily only to reduced biosecurity or increased between-farm movements. Indeed, CP culling in North Yorkshire in May was more effective than overall non-IP culling at the peak of the national epidemic, and IP culling in that area was not substantially slower than elsewhere. Thus the scale of the North Yorkshire outbreak is more attributable to the many animal movements in that area in preceding weeks—movements licensed under the false assumption that the region had escaped infection. Overall, the estimated variability in β indicates that at least as much of the variability in transmission rates throughout this epidemic is attributable to factors unrelated to culling (such as biosecurity and enforcement of movement control) as to the changes in culling policies and their implementation.

The number of discontinuous land parcels associated with a farm holding was significantly correlated with risk of FMD infection (linear slope estimate 0.11, 95% confidence limits 0.08 and 0.14), and this, together with farm-type variability, explained more than 65% of the observed geographical variation in transmission intensity. Cumbrian farms were significantly more fragmented than those in Devon or Wales, and this seems to have increased transmission (in contrast to an earlier theoretical prediction¹²), presumably as a result of a higher frequency of personnel and vehicle movement between land parcels. Given that fragmentation is a one-dimensional

summary measure of local farming environments, and given the limitations of the data used in its estimation, the magnitude of the effect seen is remarkable. If farm-type variability and fragmentation were ignored, the spatial hazard model (see Methods) overestimated R (compared with the value directly estimated from case data) in Devon by 46% and underestimated it in Cumbria by 18%. These figures were reduced to 35% and 7% by allowing for farm type, and to 24% and 1% once fragmentation was included in the analysis (assuming a linear relationship between farm susceptibility and the number of constituent discontinuous land parcels). Total land area of farms (the only other variable available for all census farms) was also examined, but as it was confounded with fragmentation and significantly less predictive of risk ($P < 0.001$ using a likelihood ratio test), it is excluded from the results presented here. Some correlation between fragmentation and animal numbers was also seen (as might be expected), but this was less significant than for land area, and was minimized by the choice of size categories made.

Figure 2d illustrates the estimated impact of the different non-IP culling programmes on disease transmission throughout the epidemic, as represented by the proportion of the contact neighbourhood of an IP (defined by the estimated spatial kernel of transmission) removed within 7 days of its slaughter. CPs represent about 16% of the contact neighbourhood, so full implementation of the CP culling policy alone would have resulted in removal of this proportion within 48 h, whereas the total effect of non-IP culling peaked in early April at 10% of the neighbourhood removed in 7 days, before dropping to less than 5% a month later. The direct impact (as compared with the cumulative effect of local farm depletion) of neighbourhood removal on transmission (measured by R) depends on the delay from IP report to culling of surrounding farms, so although prompt culling of CPs is estimated to reduce R by more than 12%, delays in culling substantially reduce this figure and thus policy efficiency.

Judged by the neighbourhood removal per farm culled within 7 days, the efficiency of policies differed significantly, with culling of CPs being three times more efficient than the 3-km or local culls, as the latter involved only sheep and typically targeted farms more distant from IPs. This implies that 1,500 fewer farms would have needed to be culled for similar overall effect had a purely CP/DC policy been adopted. After new outbreaks in early May, culling rates of CPs increased once more, achieving just over 10% neighbourhood removal within 7 days of IP slaughter.

The estimates presented in Fig. 2 allow a map of relative risk of FMD spread in different regions of Great Britain to be derived (Fig. 3a) from data on the density of livestock farms (Fig. 3c) and the degree of fragmentation of farmland (Fig. 3d) with farm-type susceptibility and infectiousness weightings (Fig. 2b). It is clear that combining variation in both farm density and contiguity much better reflects observed incidence differences (Fig. 3b) than either factor alone. There are some areas of predicted high risk not yet affected by the epidemic—the Derbyshire Dales and southwest Wales, in particular—that warrant heightened surveillance and continued vigilance in maintaining movement controls and biosecurity measures.

Translating the impact of culling on contact neighbourhoods into an estimated effect on epidemic progression requires a dynamical model of virus transmission. Here we extend an existing pair-correlation model of FMD transmission¹ to incorporate estimates of the risk factors described above (see Methods). This model, although not spatially explicit, captures key aspects of the spatial correlation structure between farms that determine the pattern of disease transmission, while being considerably less computationally intensive than a microsimulation model. The validity of this approach in describing the overall population dynamics of the FMD epidemic is demonstrated by the quality of fit achieved to the observed time series of newly infected, confirmed and slaughtered farms (Fig. 4a) obtained using the time-varying transmission

coefficient estimated from the spatial hazard model (Fig. 2c). We use this framework to examine the predicted effect on the epidemic of alternative control policies, varying both the timing and level of culling while keeping the temporal variation in the underlying transmission coefficient β of the disease the same as that shown in Fig. 2c. Implicitly, we therefore assumed in simulating these scenarios that non-culling control policies (notably biosecurity measures and movement restrictions) would have been similar through time to that observed regardless of the culling policy employed. In reality this would have been unlikely to have been the case (particularly for the scenarios in Fig. 4, where case numbers grow rapidly in June–July), but the assumption is necessary when assessing the impact of culling alone on the pattern of the epidemic.

Although IP culling alone (even if achieved within 24 h of case report from the start of the epidemic) would have significantly reduced farm culling rates in the first half of the epidemic, its effect on transmission would have been insufficient to bring the epidemic under control, particularly from May onwards (Fig. 4a and b). Figure 4 also illustrates the predicted effect of full implementation of the 24-h IP culling and 48-h CP culling targets, and of selective

improvement in either one of those policy arms. If achieved from the start of the epidemic, meeting the 24-h goal would have reduced the total epidemic size (and number of farms culled) by more than 40% up to 16 July, and implementing both policies from the start would have reduced both case numbers and farms culled by more than 60%. Full implementation from 1 April would still have reduced cases by 16% and cull numbers by 30%.

We examined the effect of regional variation in transmission (Fig. 4c and d): non-IP culling was much more critical to controlling the epidemic in Cumbria than in Devon owing to the more intense transmission there, but in both regions full implementation of the 24/48-h policy from the start of the epidemic would have resulted in more than 60% fewer cases and more than 45% fewer culled farms.

The 1967 FMD epidemic had a long and variable decay phase, which has been argued to be due in part to too-early relaxation of movement restrictions and biosecurity measures¹³. A similar picture has emerged from the current epidemic, with the decline in disease incidence being accompanied by demands from some farmers and veterinarians for a relaxation in CP culling and movement restrictions. Such frequency-dependent changes in non-culling controls

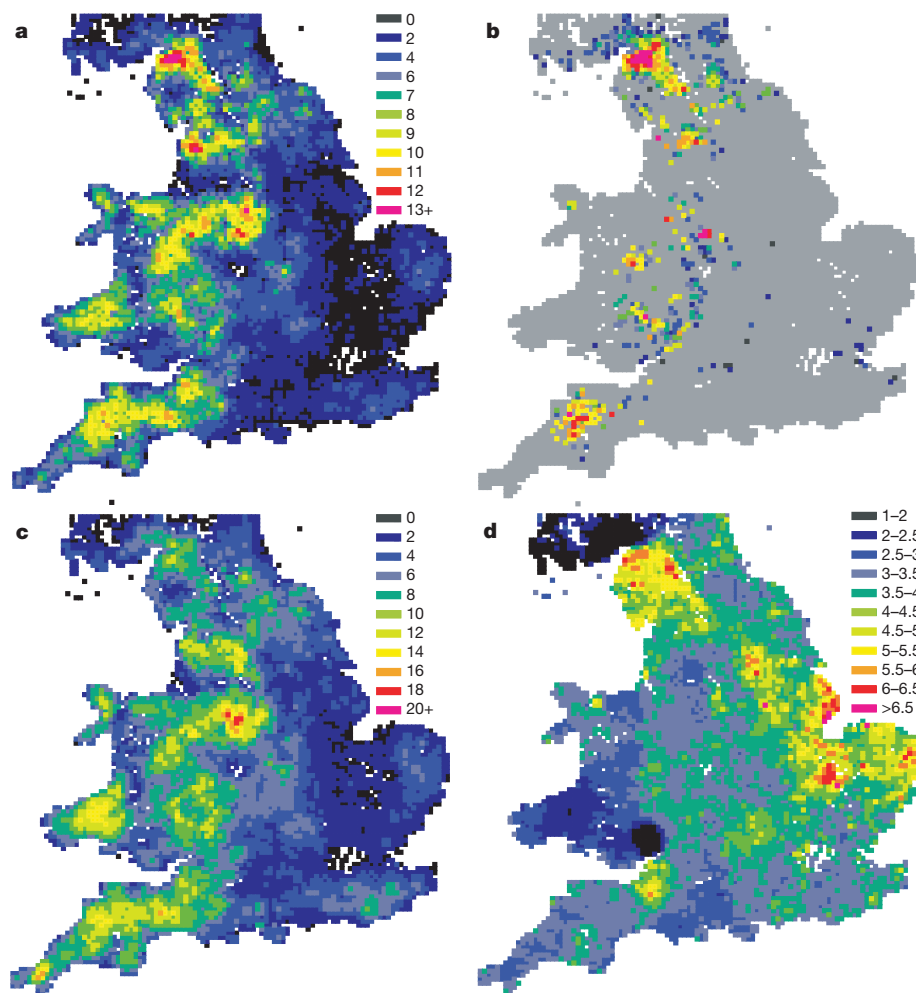


Figure 3 Areas at risk of FMD in England and Wales. Incorporation of farm fragmentation data, as well as farm density data weighted for susceptibility and infectiousness, predicts observed patterns of cases better than either factor individually. **a**, Relative transmission risk for farms averaged over 5-km squares. Transmission risk is calculated as the infectiousness of a farm to all other farms, weighted by their relative susceptibility (as determined by farm type and distance from the reference farm), because farms with increased susceptibility are more likely to become infected and thus infectious. **b**, As **a**, but

calculated solely for IPs. IPs are at higher risk than non-IPs in the same neighbourhood. **c**, Farm densities, spatially smoothed with the spatial transmission kernel and averaged over 5-km squares. **d**, The average number of disconnected land parcels in the neighbourhood of farms within a 5-km square, spatially smoothed with the spatial transmission kernel. Scotland was excluded because of inconsistencies in the land-parcel data between it and England and Wales. See Supplementary Information for details on the calculation of these statistics.

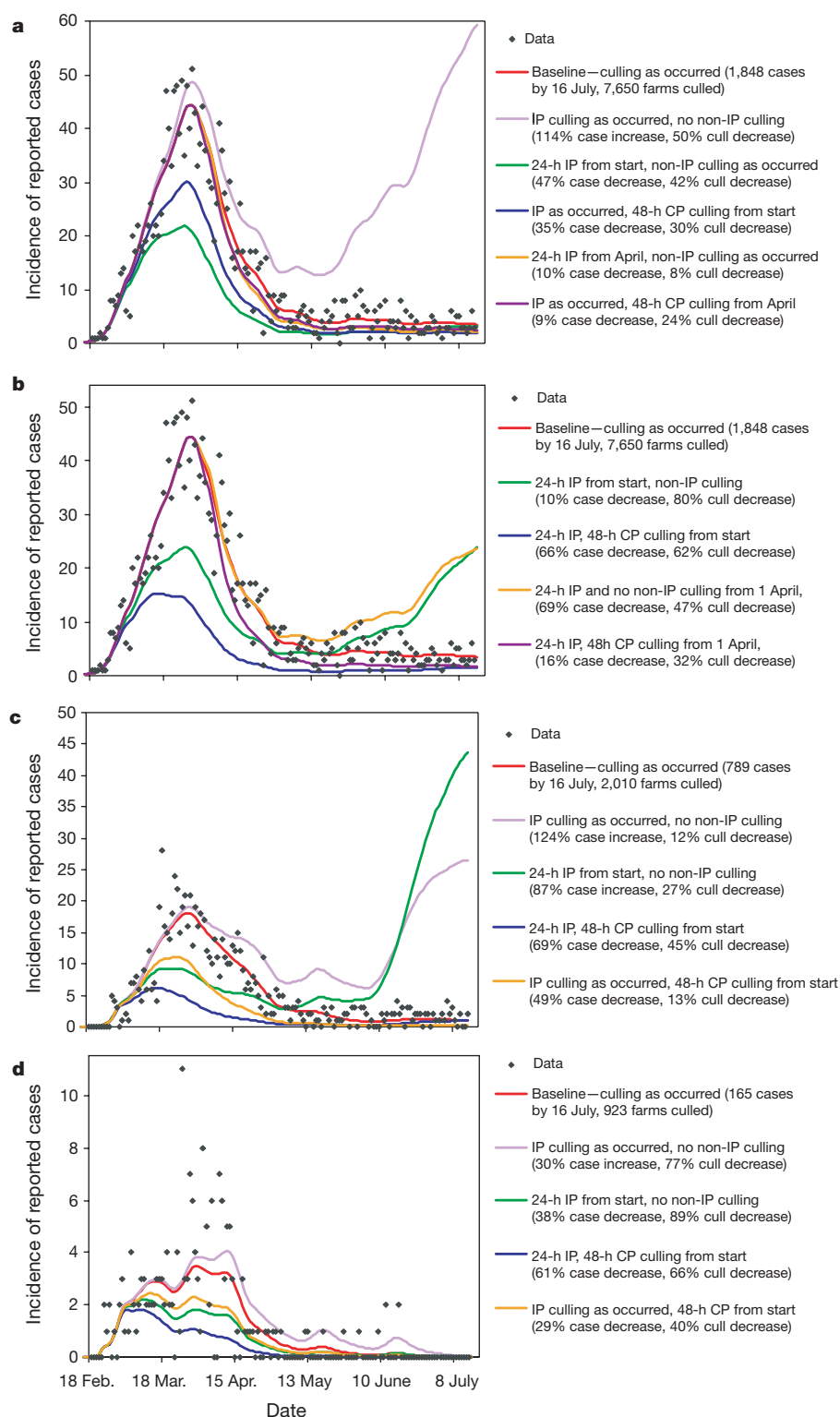


Figure 4 Simulated FMD epidemic time series. We generated the series with a pair-correlation epidemic model (see Methods), using estimates of β (Fig. 2c) and observed non-IP culling levels and timings, together with time-varying distributions of infection-to-report, report-to-confirmation and report-to-slaughter delays estimated from the FMD case data. **a, b**, Observed and predicted case incidence for all infected areas of Great Britain (45,000 farms modelled) assuming the same time-varying pattern in β , but varying the patterns of IP and non-IP culling from the observed baseline in timing and magnitude of effort. For each scenario modelled, the relative numbers of predicted cases and farms culled are given. For example, had both 24- and 48-h targets been achieved from the start of the epidemic, the model predicts 66% fewer cases and 62% fewer farms culled by 16 July. **c**, Observed and predicted case incidence for Cumbria (6,000 farms modelled)

assuming the same Cumbria-specific time-varying pattern in β for each scenario. **d**, Observed and predicted case incidence for Devon (8,000 farms modelled) assuming the same Devon-specific time-varying pattern in β for each scenario. The quality of fit of the model for Devon is similar to that for Cumbria, despite the apparent greater variability due to lower case numbers. For Cumbria and Devon, as for Great Britain, we varied culling patterns and compared resulting model predictions with the observed baseline prediction. Region-specific estimates of β were obtained from stratified output from the spatial hazard model (see Methods) fitted to the entire epidemic, whereas the region-specific delay distributions were fitted exclusively to regional data. The same estimated spatial transmission kernel was used in each case.

(for example, biosecurity and movement restrictions) have led to considerable (largely unpredictable) variability in transmission rates over past months. Past outbreaks in areas such as North Yorkshire and (more recently) Northumberland also indicate that stochastic low-probability long-range transmission events causing significant outbreaks in new areas are likely to be a central factor in determining the duration of the epidemic tail, even if such outbreaks are controlled relatively rapidly.

The unpredictability of these factors, together with the stochasticity intrinsic to the tails of epidemics, lead us to avoid making predictions of the extinction date of this epidemic here. However, our analyses have demonstrated the necessity of rapid and complete IP and CP/DC culling for disease control and eventual elimination. Additionally, the observation that changes in culling policies explain less than 50% of observed past variation in transmission rates indicates that effective movement restrictions and rigorously maintained biosecurity are equally vital. These priorities are reflected in recent government actions to enhance these controls, both in infection hotspots and through restrictions on autumnal animal movements. Such controls are particularly critical in the regions of highest transmission risk (for example, Cumbria, North Yorkshire and Northumberland) highlighted by our analysis, even where no active infection seems to be present (such as in Derbyshire and Powys). Additional outbreaks are probable before the disease is eliminated from Great Britain, so continued vigilance is also essential, both in rapid clinical diagnosis and through enhanced serological surveillance to identify any asymptomatic infection in sheep populations. □

Methods

Additional detail is provided as Supplementary Information.

Data

An integrated data set was created from DEFRA FMD-specific databases (Disease Control System and Veterinary Laboratories Agency (VLA), a database created from IACS (Integrated Administration and Control System) data by the Farming and Rural Conservation Agency on farmland fragmentation, a database extracted from the June 2000 agriculture census¹⁴, and a database describing animals culled under the Livestock Welfare (Disposal) Scheme. The data from each of these sources were linked together using the county-parish-holding (CPH) unique farm identifier to form an integrated database. The key variables analysed for each farm were location, number of cattle, sheep and pigs, and number of land parcels, with additional variables analysed for both IP and non-IP culled farms: dates of FMD report, confirmation and culling for IPs, and date and reason for culling for non-IP culled farms. In the absence of accurate or comprehensive adjacency data, the contiguous premises (CPs) of each IP were identified from approximated farm boundaries obtained with the Quickhull algorithm¹⁵ for constructing Voronoi diagrams from the point locations of farms reported in the June 2000 agricultural census¹⁴.

Spatial infection model and parameter estimation

Owing to the lack of certainty in the source of most farms' infections, source farms are attributed probabilistically on the basis of their relative infectiousness (to infected farm i) on the day farm i was infected. Farms were classified into infectiousness and susceptibility on the basis of farm type (defined by the number of animals of each species), and susceptibility was assumed to be a function of farm fragmentation. The probability that infected farm i was infected by farm j is proportional to the infectiousness of farm j relative to the infectiousness of other potential source farms (a function of their farm type, distance from farm i , and potentially the time since infection of the source farm). The sum of the proportions of infections attributed to farm j is the farm-specific estimate of R , which, corrected for neighbourhood depletion, gives the farm-specific R_0^i estimate. The transmission coefficient is estimated as the ratio of the number of infections occurring on a particular day, corrected for neighbourhood depletion, to the estimated total relative infectiousness of infected farms on that day.

The form of the dependence of infectiousness on the time since infection is assumed to be constant from the day after a farm is infected until the day its animals are culled. Other parameters are estimated iteratively, after initially assuming all farm types are equally infectious and susceptible, by adjusting parameters (see Supplementary Information) such that the observed and predicted number of infections in each class (infectiousness, susceptibility or distance from source farm) were equal. The effect of farm fragmentation is estimated by assuming susceptibility was linearly proportional to fragmentation and fitting the slope parameter such that the average number of fragments in the farms expected to be infected by IPs over the course of the epidemic equalled the average number of fragments in observed IPs (see Supplementary Information).

On the basis of these parameter estimates, the likelihood of the observed epidemic was calculated based on the farm type, fragmentation and location of each farm in the country.

For each day, expected infection probabilities are obtained from the estimated hazards scaled to sum to the observed number of infections across the country that day. Univariate likelihood profiles with respect to each of the estimated parameters were generated to calculate confidence bounds on parameters. These profiles demonstrated that the parameter estimation techniques outlined above gave results very close to the maximum likelihood estimates, although being much less computationally intensive than maximum likelihood estimation using generalized nonlinear multidimensional optimization methods. Likelihood-based hypothesis tests and confidence intervals were obtained using the result that twice the difference in the full log likelihood is chi-squared distributed with degrees of freedom equal to the number of constrained parameters. This spatially explicit hazard model can also be used to calculate the expected numbers of cases generated by IPs by region for comparison with the estimates of R made directly from the case data as described above.

Pair-based transmission model

We previously developed a pair-based model of FMD transmission and fitted it to the current FMD epidemic (1) allowing for the distribution of delays between infection, confirmation and culling of IPs; (2) assuming that the transmission coefficient was constant except for a drop when movement restrictions were imposed; and (3) using the spatial kernel estimated from the distances between infected farms and their sources as identified by DEFRA staff (where available)¹. Using the same model framework (in which spatially localized disease spread is modelled by tracking the time evolution of disease state correlations between neighbouring farms), we extended the model to allow for variable infectiousness and susceptibility of farms and a time-varying transmission coefficient (estimated as described above, and scaled to match the total observed number of confirmed cases). The only parameter estimated from the fit of the transmission model to the observed time series is the date of the first infection. Species and size variation between farms was not explicitly modelled for reasons of computational simplicity, but was taken into account in the transmission coefficient, culling hazard and neighbourhood size estimates used in the model. Estimates of two key model parameters¹—neighbourhood size (25.9) and connectedness (0.05)—were derived from census data using the spatial kernel estimates presented above.

We modelled the time-dependent probability of farms in the vicinity of IPs being culled in a non-IP culling scheme using competing time-varying hazards to model both the amount and timing of culling, such that

$$1 - \int_{t_0}^t \lambda_1(t') \exp \left[- \int_{t_0}^{t'} \lambda_1(t'') + \lambda_2(t'') dt'' \right] dt'$$

is the proportion of the neighbourhood present at time t_0 that has not been culled as a non-IP by time t . The functions $\lambda_1(t)$ and $\lambda_2(t)$ were fitted (with splines) to the observed pattern of farm neighbourhood removal, stratified by both chronological time and the delay between slaughter of an IP and culling of neighbouring farms. The accuracy of the culling model and of the pair-correlation approximation is demonstrated by the close match between the model output and observed disease incidence and culling data, both through time and cumulatively.

Received 10 August; accepted 21 September 2001.

1. Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. The foot-and-mouth epidemic in Great Britain: Pattern of spread and impact of interventions. *Science* **292**, 1155–1160 (2001); published online 12 April 2001 (10.1126/science.1061020).
2. Woolhouse, M. *et al.* Epidemiology—Foot-and-mouth disease under control in the UK. *Nature* **411**, 258–259 (2001).
3. Anonymous. Lessons from an epidemic. *Nature* **411**, 977 (2001).
4. Anonymous. What is wrong with the current policy? *Guardian Unlimited* [online] (cited 20 March 2001) (<http://www.guardian.co.uk/Archive/Article/0,4273,4154893,00.htm>) (2001).
5. Highfield, R. Delays 'doubled' foot and mouth toll. *Telegraph* (2 July 2001). Available online at (<http://portal.telegraph.co.uk/news/main.jhtml?xml=%2Fnews%2F2001%2F07%2F02%2Fnm02.xml>).
6. Donaldson, A. I., Alexanderson, S., Sorensen, J. H. & Mikkelsen, T. Relative risks of the uncontrollable (airborne) spread of FMD by different species. *Vet. Rec.* **148**, 602–604 (2001).
7. UK Ministry of Agriculture, Fisheries and Food. Epidemiological forecasts: a comparison between models. Press release (4 April 2001).
8. UK Ministry of Agriculture, Fisheries and Food. Foot and Mouth Disease Nick Brown's Statement to House: Thursday 26 April 2001. Press release, available online at (<http://www.defra.gov.uk/news/newsrel/2001/010426b.htm>).
9. Anonymous. Foot-and-mouth disease: MAFF refines its policy on the contiguous cull. *Vet. Rec.* **148**, 550–552 (2001).
10. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans* (Oxford Univ. Press, Oxford, 1991).
11. McColl, K. A., Westbury, H. A., Kitching, R. P. & Lewis, V. M. The persistence of foot-and-mouth disease virus on wool. *Aust. Vet. J.* **72**, 286–292 (1995).
12. Kao, R. R. Landscape fragmentation and foot-and-mouth disease transmission. *Vet. Rec.* **148**, 746–747 (2001).
13. *Northumberland Report: The Report of the Committee of Inquiry on Foot and Mouth Disease* (HMSO, London, 1968).
14. UK Ministry of Agriculture, Fisheries and Food, the National Assembly for Wales Agriculture Department, and the Scottish Executive Rural Affairs Department. *June 2000 Agricultural and Horticultural Census* Crown copyright (2001).
15. Barber, C. B., Dobkin, D. P. & Huhdanpaa, H. T. The Quickhull algorithm for convex hulls. *ACM Trans. Math. Software* **22**, 469–483 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

For access to data we thank P. Helm for facilitating data provision; J. Wilesmith for provision of the VLA FMD case data; A. Wray, T. Sharp and M. Horn for providing census and location data; B. Froggatt, S. Newell and P. Swanton for analysing IACS data and providing a database on farmland fragmentation; F. Mallin and J. Moodey for provision of the LW(D)S (Livestock Welfare (Disposal) Scheme) data; and J. Bell and C. Cox for providing CTS (Cattle Tracing System) data. We thank members of the Chief Scientist's advisory group on FMD for discussions. N.M.F. thanks the Royal Society and the Howard Hughes Medical Institute, and C.A.D. and R.M.A. thank the Wellcome Trust for research funding.

Correspondence and requests for materials should be addressed to N.M.F. (e-mail: neil.ferguson@ic.ac.uk).

erratum

The timing of the last deglaciation in North Atlantic climate records

Claire Waelbroeck, Jean-Claude Duplessy, Elisabeth Michel, Laurent Labeyrie, Didier Paillard & Josette Duprat

Nature **412**, 724–727 (2001).

In this Letter, the following line should have appeared before the reference list: "Received 10 November 2000; accepted 5 July 2001."

Control of conformational and interpolymer effects in conjugated polymers

J. Kim & M. Swager

Nature **411**, 1030–1034 (2001).

In Fig. 2a the pressure–area isotherms for polymers 2 and 3 were incorrectly labelled as 'Polymer 3' and 'Polymer 2', respectively.

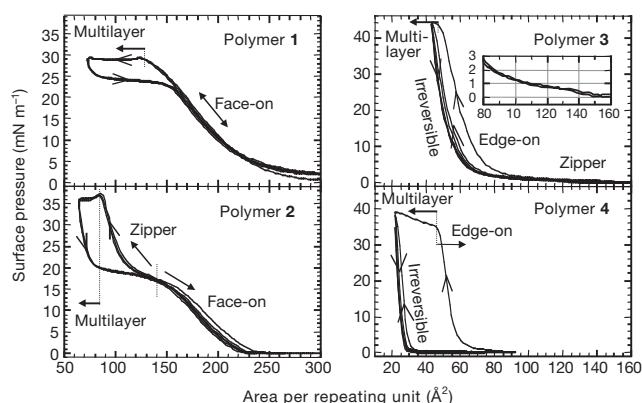


Figure 2a

Partitioning selection and complementarity in biodiversity experiments

Michel Loreau & Andy Hector

Nature **412**, 72–76 (2001).

In Figures 2 and 3, some regression lines were very faint or missing. The corrected panels are reproduced below. All panels in both figures should also have shown the null hypothesis as a horizontal dashed line through zero.

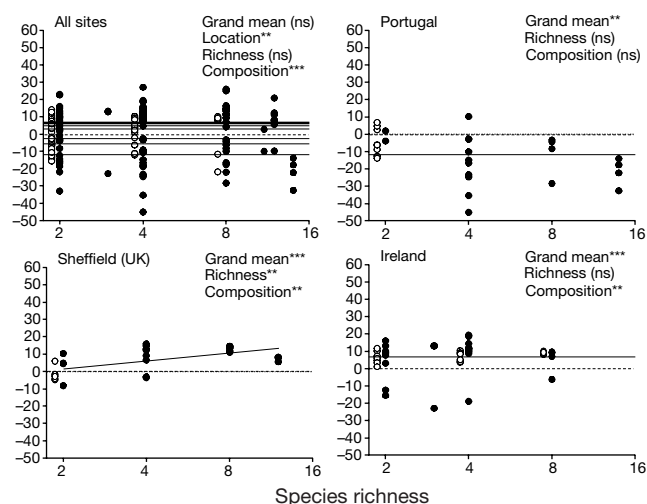


Figure 2

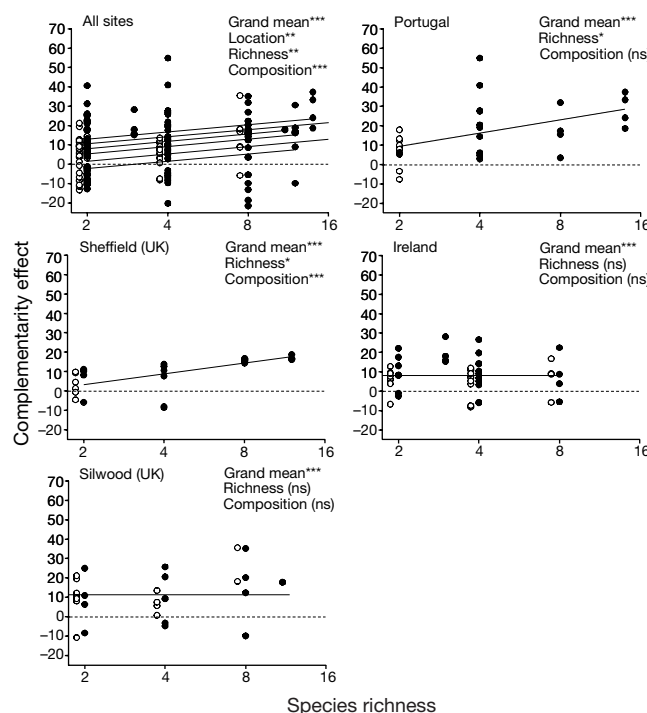


Figure 3

High-throughput screening

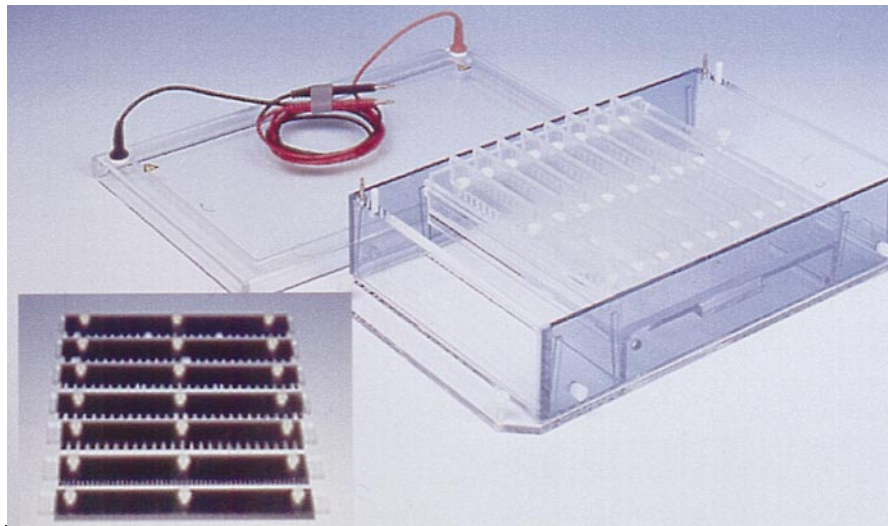
Assay kits, plates, probes and rotors for high throughput.

Submarine electrophoresis

C.B.S. Scientific www.cbsscientific.com

Prepared to go to any length

Designed for rapid screening of cloning products, restriction fragments, synthetic oligos and PCR products, the new horizontal gel electrophoresis systems from C.B.S. are available in three lengths and feature a microtitre format for quick multichannel pipetting. The trays have height-adjustable multiple comb slots, with each comb capable of handling 25 to 102 samples simultaneously. Combs are made of polycarbonate or, for more rapid heat dissipation, Teflon-coated aluminium, particularly suitable for use with low percentage agarose gels. The systems are available with 26×14 cm, 26×26 cm or 26×40 cm gel bed dimensions. They incorporate a cooling chamber and accommodate a variety of comb/path lengths. All combs have microtitre spacing and are available in thicknesses ranging from 0.75 to 2.0 mm with 102, 51, 28, 27, 26 and 25 wells. One comb is supplied as standard equipment with each unit. Other standard features include a UV transparent gel tray for direct reading of ethidium bromide stains, tapered baffle system for submarine gel formation and buffer and coolant circulation ports.



Horizontal hold; rapid gel; electrophoresis from C.B.S. Scientific.

1,536 well plates

Nalge Nunc www.nalgenunc.com

Flatter by design

Nalge Nunc's 1,536 well plates have a total well capacity of 13 μ l and are available in clear, black and white, offering flexibility in assay and detection methods. The plate design ensures optimal flatness and allows barcode labelling.

I κ B- α pSer 32

PanVera Corporation www.panvera.com

Serine kinase assay kit

The fluorescence polarization-based Core-HTS serine kinase assay kit contains the materials necessary to quantify the products of a kinase reaction in which the phosphorylated substrate resembles the amino acid sequence around phosphoserine 32 of I κ B- α . Phosphorylation of I κ B- α at serines 32 and 36 results in the activation and nuclear translocation of NF- κ B. This phosphorylation is activated by a diverse group of extracellular signals, including virus infection and oxidative stress, and is a key regulatory point in the initiation of immune or inflammatory responses. The principle behind the assay is competition: a fluorescent phosphopeptide tracer and the non-fluorescent phosphopeptides generated during a protein kinase reaction compete for

binding to an antibody specific for the phosphorylated Ser 32 on the phosphorylated I κ B- α peptide. In a reaction mixture containing no phosphopeptide product, the fluorescent tracer is bound by the antibody and the emission signal is polarized. But in a reaction mixture containing phosphopeptide product, the fluorescent tracer is displaced from the antibody and the emission signal becomes depolarized. The decrease in polarization values correlates with enzyme activity/inhibition. The kit is designed for use in high-throughput drug discovery screening and is available in 100, 1,000 and bulk assay sizes.

CELISA

Biowhittaker www.cambrex.com

Cell-based efficacy and toxicity assay

CELISA is a technology for cell-based high-throughput screening technology using time-resolved fluorophore-conjugated probes and filtration technology to perform cell-based screens of therapeutic targets. Haematopoietic stem cells are the initial target. These cells originate in the bone marrow and differentiate into red or white blood cells or platelets through a process called haematopoiesis. The CELISA assay can be used to evaluate the effect of drug candidates on the relative production of different cell lineages and to ensure that the drug is not toxic to bone marrow.

Tissue microarrays

InnoGenex www.innogenex.com

High-throughput screening of gene expression

These new microarrays are histologically characterized slides from diseased and normal

human tissues, mouse normal tissues, and rat normal tissues. They can be used in high-throughput screening of *in situ* gene expression for target validation. The microarrays can also be used to identify novel antibody markers, in animal model studies, and in other applications where traditional tissue sections would be used. Low-density arrays contain 20–50 elements per slide, while the high-density arrays have 100–200 elements per slide. All donor tissue blocks and final tissue arrays are validated for use in immunohistochemistry and *in situ* hybridization assays.

Roboocyte

Ala Scientific Instruments

www.alascience.com

Smart operator

Roboocyte was developed to provide the pharmaceutical industry and academic researchers with an automated system for recording ion currents in oocytes of *Xenopus laevis* using the two-electrode-voltage-clamp (TEVC) technique in a 96-well dish. The system offers high-throughput sequential injection of cDNA/ mRNA into oocytes and subsequent TEVC recording and can operate on a 24-hour basis without supervision. Injection of cDNA/ mRNA, TEVC, in-between wash of the cells and perfusion paradigms are all performed automatically. Applications include expression screening, drug screening, including medium or high throughput of agonists, antagonists and co-agonists, expression cloning, screening of mutated or chimeric receptors or ion channels and generation of dose-response curves on one or several oocytes.

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:

Leonie Welss (4954)

European Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andrew Douglas (4975)

Holland/ Italy: Nevin Bayoumi (4978)

Scandinavia: Silje Opstrup (4994)

Spain/ Portugal: Leonie Welss (4954)

Production Manager:

Billie Franklin
To send all materials use
London address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel + 33 (0) 1 43 20 16 51
Fax + 33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mails: p.phelan@nature.com
k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Corissa Salzman

Japan Head Office, Tokyo

MG Ichigaya Building (6F),
19-1 Haraikatsunachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.cowan@naturejp.com
Japan Manager: Kate Cowan

naturejobs

nature.com

Shaking off genome fatigue

In the wake of the hoopla surrounding the publication of the rough draft of the human genome, the potential emergence of 'genome fatigue' in the media, and even among scientists, is understandable. The genome remains a work in progress, and has not yet produced any of the promised medical panaceas. Many of the jobs that policy-makers said would be in demand as a result of the stepped-up effort — production, finishing and annotation — seem like old news and, although important, are hardly glamorous.

And once the obvious organisms — mouse, chimpanzee, zebrafish — are done, then what? It's hard to imagine debates about the utility of sequencing the duck-billed platypus as opposed to the spotted owl generating much passion outside limited circles or leading to incredible new job opportunities.

Perhaps one reason for genome fatigue, beyond simple fallout from the hype, is that discussions about jobs in genomics focus too much on sequence production. It seems almost churlish to cavil at this: the increase in throughput and the rapid rise in the number of genomes in the sequencing pipeline are nothing short of amazing. It is, perhaps, even more astounding that many take this progress for granted.

But now that the hype about the sequence has died down, the genome itself is becoming intellectually interesting again. Instead of merely devising ever-faster ways to produce data, scientists are thinking about creative ways to use the existing information. A haplotype-mapping project aimed at producing a map of complex genetic variations is one idea (see *Naturejobs*, page 4). Such projects will create jobs that demand sophisticated sets of skills.

So the draft human genome should not generate complacency but continue to re-energize scientific opportunity.

Paul Smaglik
Naturejobs editor



Contents

SPECIAL REPORT

Demand for population geneticists grows p4

WWW.NATURE.COM/ NATUREJOBS

Career centre
Information on the scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Population geneticists are in short supply, as the need to translate large data sets into disease-susceptibility traits grows, says Eugene Russo

Genetics is in a state of the many relying on the few. The many are the scientists proficient at producing genomics data. The few are the experts capable of interpreting the relevance of those data across populations. As the quantity of genetics data increases, it is becoming apparent to the many that the few are not enough.

A select subset of that few, population geneticists (or, as they are sometimes now more accurately called, molecular population geneticists) — researchers with not only genetics, but also mathematics and statistical know-how — are in increasingly high demand. Their role is becoming more important as large disease-susceptibility gene projects such as the haplotype-map project (see next page) are being proposed.

"There are way more jobs than people to fill them," says Michael Boehnke, professor of biostatistics and director of the Center for Statistical Genetics at the University of Michigan. Jobs for population and statistical geneticists are so common, says Boehnke, that qualified PhD students routinely skip postdoctoral study to accept immediate faculty

appointments. Meanwhile, molecular biologists often struggle to find appointments after doing several postdocs.

Boehnke — who himself started as a mathematician before parlaying his skills into a 'biomathematics' programme at the University of California, Los Angeles — is the sole recipient of the only National Human Genome Research Institute (NHGRI) training grant that specifically emphasizes population genetics. The programme currently funds ten pre- and postdoctoral students. Boehnke, whose own centre has had problems finding permanent staff in these fields, recalls how he recently advertised the impending availability of two of his students to 15 colleagues at

companies and universities across the country. Within two days, he received eight e-mails expressing interest.

The current abundance of interest contrasts sharply with the field's early days. Population geneticists were once limited largely to the realms of maize and fruit flies. "Population genetics has been this kind of esoteric field," said Pennsylvania State University population geneticist Andrew Clark. "It got relegated to the back of journals like *Genetics*."

Eric Lander, director of the Whitehead Institute in Cambridge, Massachusetts, agrees. But he senses a change. "Now we're realizing that the human population, because it is so well-studied medically, can be a place for primary discovery."

THEORETICAL HISTORY

Population-genetics pioneers such as Ronald Fisher, J. B. S. Haldane and Sewell Wright dealt primarily in theoretical terms, then tested their ideas with plant and animal models. But today's population geneticist has something those experts of the 1920s and 1930s lacked: data. Loads of them.

The Human Genome Project and projects born of it continue to churn out vast amounts of sequence data. That data flood has also produced disease-susceptibility projects such as the search for genetic markers in its wake. The largest effort so far has been the search for single nucleotide polymorphisms (SNPs) — variations in only one base pair.

The public-private SNP Consortium and private companies such as Celera Genomics have produced several million SNPs in the past two years, in the hope of identifying significant markers of disease. But, while SNPs are relatively easy to find, their use in tracking down disease susceptibility may be limited.

A new project under discussion could make these data more useful by mapping sets of SNPs that are collectively inherited in close proximity to one another. But before tens of millions of dollars are devoted to mapping these haplotypes, academic and industrial scientists must first determine the project's feasibility and utility.

Formal discussions began this summer. But the input of population geneticists may determine whether and how the project will proceed. Before they decide, they must discuss the prevalence of haplotype structures in the genome and the degree to which one

Potential partners

Early discussions have hinted at the possibility of a hap-map project public-private consortium, a model recently adopted by the SNP Consortium. Possible members include:

The SNP Consortium
<http://snp.cshl.org>
 Whitehead Institute
<http://www.wi.mit.edu>
 Sanger Center
<http://www.sanger.ac.uk>

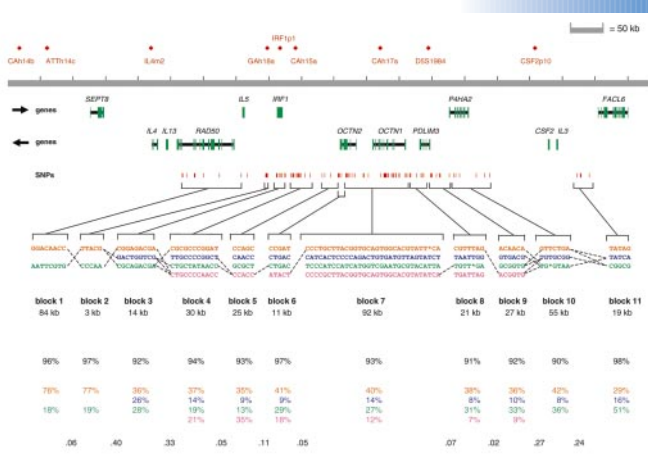
US National Human Genome Research Institute
<http://www.nhgri.nih.gov>

Genome Canada
<http://www.genomecanada.ca>

University of Tokyo's Human Genome Center
<http://www.hgc.ims.u-tokyo.ac.jp>

Orchid Biosciences
<http://www.orchid.com>

Sequenom Industrial Genomics
<http://www.sequenom.com>



population's haplotype map will apply to another.

The need for population geneticists already extends beyond any prospective haplotype map project. These scientists are essential to everything from positional cloning of complex traits to tracking species' origins.

The NHGRI is doing a few things to encourage more scientists to enter that field, in addition to the training-grant programme that helped Boehnke.

Although it is not petitioning the population-genetics community as a whole to submit grants, the institute is encouraging scientists with mathematics, engineering and computer backgrounds who have biology skills to apply for career development grants.

According to NHGRI deputy director Elke Jordan, in recent years the NHGRI, and other institutes including the National Institute of Mental Health and the National Institute of General Medical Sciences, have awarded many more research grants to population geneticists and the like. "Their role is changing, as they are now moving much more into mainstream biomedical research," says Jordan. "Molecular-biological and population-biological approaches are merging to complement and enhance each other."

ENCOURAGING CROSS-TALK

Mike Boyce-Jacino of Orchid BioSciences, a biotechnology firm based in Princeton, New Jersey, and a potential partner in the haplotype-map project (see "Potential partners", page 4) sees a growing need for population geneticists in both industry and academia who can determine the best populations to study, the necessary density of markers, and how to interpret results. "There is a shortage of people who can both develop the algorithms for analysing the data and also actually interpret the results," he said. "There aren't too many people that understand the link between the genetics side and statistics."

A course or two in statistics will not suffice, says Andrew Clark. Nor will the packaged statistical software programs used by many molecular biologists.

One obstacle to better understanding of complex human diseases is that relatively few medical geneticists know about evolution, and relatively few population geneticists understand the intricacies of disease epidemiology. "The problem is there's not enough cross-talk," said Columbia University genome

centre medical geneticist Joseph Terwilliger.

In order to understand variation in populations, scientists must also understand history, evolution and selection. "And selection is not something that matters in the typical molecular biology lab day to day," says Lander.

Eugene Russo is a freelance science writer based in Takoma Park, Maryland.

Obstacles for hap-map project

A haplotype-map project promises to speed up disease-gene discovery, but it faces major hurdles.

SNPs are simple genetic variations loosely associated with disease. Haplotypes are collections of these variations that occur in close enough proximity to be inherited together.

Rather than genotyping SNPs one at a time, the project would generate a haplotype map of the entire genome. Much of the SNP information within a haplotype is redundant; scientists may therefore require only a few SNPs within a haplotype in order to link it with a disease.

Findings reported in *Nature* in May (Reich, D. E. *et al.*, *Nature* 411, 199–204; 2001) and in the October issue of *Nature Genetics* suggest that haplotype structures are consistent

across the genome. But researchers must still identify the necessary density of markers and analytical tools and the nature of the population samples to be used.

Some critics argue that doing the project across the entire genome is a waste of time and money.

The map's usefulness across populations hinges on the 'common disease-common variant' hypothesis, assuming that most variants for disease are in all populations. But, asks Lisa Brooks of the NHGRI, "How different are populations in their haplotype structure?"

Mike Boyce-Jacino of Orchid BioSciences voices ethical concerns: "When you go to haplotype level, people are more nervous that you're associating gene function almost on a population-specific basis," he warns. E.R.

Top left: Eric Lander

Above: SNP

Elke Jordan



Mike Boyce-Jacino

